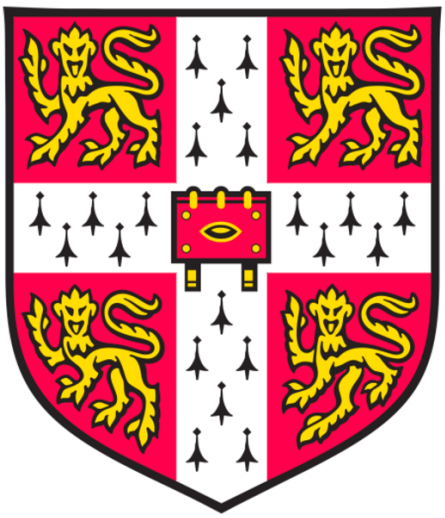
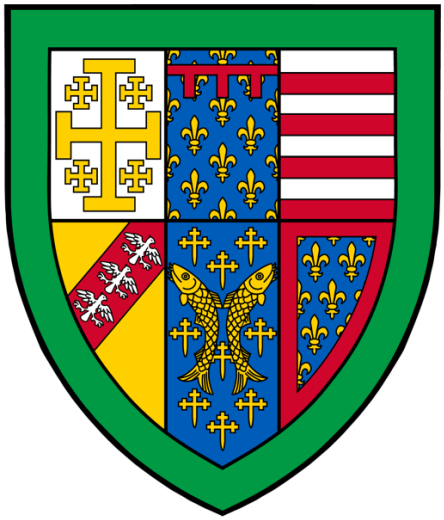
Deciphering pre-leukaemic dysregulation caused by mutant C/EBPα at the level of the single cell





Moosa Rashid Qureshi

Department of Haematology

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Queens' College

August 2019

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Acknowledgments and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Acknowledgments and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Acknowledgments and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee

Moosa Qureshi

August 2019

Abstract

Deciphering pre-leukaemic dysregulation caused by mutant C/EBPα at the level of the single cell

Moosa Rashid Qureshi

C/EBPα is an ideal candidate to explore pre-leukaemic dysregulation of transcriptional networks because it is a significant player in normal myelopoiesis, it interacts with other transcription factors (TFs) in lineage specification, and the *CEBPA* gene has been identified as an early mutation in acute myeloid leukaemia (AML). N321D is a mutation affecting the leucine zipper domain of C/EBPα which has previously generated an aggressive AML phenotype in a mouse model.

The first objective of my thesis was to construct a novel *in vitro* cellular model to recapitulate pre-leukaemic changes associated with the *CEBPA* N321D mutation. The conditionally immortalised murine Hoxb8-FL cell line can differentiate into relatively homogenous populations of functional myeloid, lymphoid and dendritic cells in appropriate cell culture conditions. These cells were transduced with retroviral vectors which expressed either no transgene or "Empty Vector" (EV), the *CEBPA* wildtype (WT) or N321D mutation. *CEBPA* N321D-transduced cells had immature morphological appearance in Flt3L differentiation media, proliferated indefinitely, and were characterised by a plasmacytoid dendritic cell (pDC)-like immunophenotype. These cells were capable of long-term engraftment in a mouse model but did not recapitulate leukaemia for several months after transplantation.

The second objective was to characterise the transcriptional deregulation associated with the phenotypes demonstrated by EV, *CEBPA* WT and N321D-expressing cells in the presence of Flt3L. Single cell RNA-seq (scRNA-seq) confirmed phenotypic observations that *CEBPA* N321D-transduced cells had a pDC precursor-like gene expression signature when placed in Flt3L conditions of differentiation. Single cell analysis also identified transcriptional heterogeneity in the *CEBPA* N321D compartment at single cell resolution.

The third objective was to use ChIP-seq to investigate the effect of the N321D mutation on C/EBPα's TF binding profile, and correlate this with scRNA-seq data to identify a set of candidate genes (*NOTCH2*, *JAK2*, *SIRPA*, *FOS*, and others) where gene expression changes were associated with direct binding events by the mutant TF.

In addition to characterising the transcriptional effects of the *CEBPA* mutation, a fourth objective was to update and expand the Göttgens group's CODEX database of TF-binding and histone modification data for human blood cells, increasing annotation of the genome to 14.83%. This online database now provides significant coverage of the non-coding portion of the genome, and is therefore potentially of interest to the wider haematology research community.

Acknowledgements

Firstly, I would like to thank my supervisor Bertie Göttgens for his invaluable mentorship and support during my PhD, particularly for providing vision at those definitive points when I could not myself see a clear path forwards. It has been a real privilege to work under his supervision.

I am indebted to everyone in the Göttgens lab for providing a fantastic environment of academic endeavour, but in particular I would like to thank Fernando Calero-Nieto for providing me with practical help and advice throughout my scientific apprenticeship, and for teaching me the importance of scientific rigour. In addition, I would like to thank Sarah Kinston and Iwo Kucinski for their practical help with cloning and single cell work, and the Flow Cytometry Team at the CIMR for helping me to develop a workable flow panel and for cell sorting. From the Göttgens bioinformatics team, I would like to thank Evangelia Diamanti, Fiona Hamey, Xiaonan Wang, and Wajid Jawaid for helping me with R-code, and especially Rebecca Hannah who helped enormously with processing my ChIP-seq data and my work on the CODEX database. It would be greatly remiss of me not to acknowledge the outstanding administrative support I received from Martin Dawes. Regarding collaboration, I would like to thank the Huntly group and in particular George Giotopoulos for his help and expertise with *in vivo* work.

Thirdly, I thank the Cambridge Cancer Centre for selecting me for this PhD, and Cancer Research UK and the Addenbrooke's Charitable Trust for supporting my research. In addition, I would like to thank the Charles Slater Fund and Queens' College for funding me to present my work at international conferences.

I would also like to thank my father who continues to inspire me, my mother whose memory sustains me, and all my extended family. Finally, I express my gratitude to Sara Khalil Ebrahim Ali for providing love, intellectual stimulation, asking me questions which I can never answer, and for simply putting up with me.

Abbreviations

AML	Acute Myeloid Leukaemia
APC	Antigen Presenting Cell
APML	Acute Promyelocytic Leukaemia
ASAP	Automated Single-cell Analysis Pipeline
BMDC	Bone Marrow derived Dendritic Cell
BMS	Bone Marrow Stroma
BR	Basic Region
bZIP	basic-region leucine Zipper
C/EBP	CCAAT/Enhancer Binding Protein
cDC	Conventional Dendritic Cell
cDNA	Complementary DNA
CDP	Common Dendritic cell Progenitor
CHIP	Clonal Haematopoiesis of Indeterminate Potential
ChIP	Chromatin Immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation Sequencing
CLP	Common Lymphoid Progenitor
CMGA	Complex Molecular Genetic Abnormalities
CML	Chronic Myeloid Leukaemia
CMP	Common Myeloid Progenitor
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CyTOF	Cytometry by Time Of Flight
DC	Dendritic Cell
DGE	Differential Gene Expression
dm	double mutated
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphates
dT	Deoxythymine
EDTA	Ethylenediaminetetraacetic Acid
ENCODE	ENCyclopaedia Of DNA Elements
ERCC	External RNA Controls Consortium

EV	Empty Vector
FAB	French-American-British (AML Classification System)
FACS	Fluorescence-Activated Cell Sorting
FBS	Foetal Bovine Serum
G-CSF	Granulocyte-Colony Stimulating Factor
GEO	Gene Expression Omnibus
GFP	Green Fluorescent Protein
GM-CSF	Granulocyte Macrophage-Colony Stimulating Factor
GMP	Granulocyte Macrophage Progenitor
GO	Gene Ontology
GRN	Gene Regulatory Network
GWAS	Genome-wide Association Study
HEPES	4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid
HPC-7	Haematopoietic precursor cell-7
HSC	Haematopoietic Stem Cell
HSPC	Haematopoietic Stem and Progenitor Cells
IFN	Interferon
IL-3	Interleukin 3
ImmGen	Immunological Genome
IRES	Internal Ribosome Entry Site
LIC	Leukaemia Initiating Cell
LMPP	Lymphoid-Primed Multipotent Progenitor
LSK	Lin^{-} $Sca\text{-}1^{+}$ $c\text{-}Kit^{+}$
LT-HSC	Long-Term HSC
LZ	Leucine Zipper
MACS	Model-based Analysis of ChIP-seq
mChrom1	mouse Chromosome 1
M-CSF	Macrophage-Colony Stimulating Factor
MDP	Monocyte/Macrophage Dendritic Cell Progenitor
MDS	Myelodysplastic Syndrome
MEP	Megakaryocyte Erythrocyte Progenitor
MHC	Major Histocompatibility Complex
MLL	Mixed Lineage Leukaemia
MLP	Multi-Lymphoid Progenitor
MPP	Multipotent Progenitor

mRNA	messenger RNA
NOD SCID	Non-Diabetic with Severe Combined Immunodeficiency Disease
NGS	Next Generation Sequencing
NSG	NOD SCID Gamma
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
pDC	Plasmacytoid Dendritic Cell
PBS	Phosphate Buffered Saline
qPCR	Quantitative PCR
RNA	Ribonucleic acid
RPMI	Roswell Park Memorial Institute Medium
sc	single cell
SCF	Stem Cell Factor
Scl	Stem cell leukaemia
SD	Sequencing Depth
shRNA	short hairpin RNA
smFISH	single-molecule Fluorescence *In Situ* Hybridisation
SNP	Single Nucleotide Polymorphism
SNR	Signal-to-Noise Ratio
ST-HSC	Short-term HSC
TAD	Transactivation Domain
TAE	Tris/Acetic Acid/EDTA Buffer
TBE	Tris/Borate/EDTA Buffer
TBS	Tris Buffered Saline
TE	Tris/EDTA
TF	Transcription Factor
t-SNE	t-Distributed Stochastic Neighbour Embedding
TSS	Transcription Start Site
UCSC	University of California Santa Cruz (Genome Browser)
UMI	Unique Molecular Identifier
URE	Upstream Regulatory Element
WGS	Whole Genome Screening
WT	Wildtype

Contents

List of Figures

List of Tables

List of Figure and Tables in Appendix

1. Introduction

My PhD thesis undertakes single cell analysis of pre-leukaemic transcriptional dysregulation caused by *CEBPA* N321D, and describes a dendritic cell phenotype following expression of this mutation during early haematopoiesis. In this introductory chapter, I first present current paradigms of normal and malignant haematopoiesis, focusing on early progenitors and leukaemia initiation, and placing my cellular model and dendritic cells within this framework. I then present a general synopsis of the transcriptional control of haematopoiesis, before undertaking a detailed examination of C/EBPα's role in that process. Finally, I give a brief overview of how single cell technology is changing current paradigms of haematopoiesis and its transcriptional regulation.

1.1 Normal haematopoiesis

1.1.1 Haematopoietic stem cells and progenitor populations

Haematopoetic stem cells (HSCs) are defined by dual properties of multipotency and self-renewal as demonstrated by serial transplantation experiments, whereas haematopoietic progenitor cells are characterised by lineage restriction and limited self-renewal capacity in irradiated mouse models (Weissman *et al*, 2001, Oguro *et al*, 2013). This paradigm of normal haematopoiesis was heralded by Till and McCulloch's seminal finding that rescuing irradiated mice with bone marrow transplantation led to splenic "bumps" capable of reconstituting heterogeneous haematopoietic lineages (Till *et al*, 1961), and experimental proof of clonality in these splenic colonies (Becker *et al*, 1963). Over the following decades, it was established that normal haematopoiesis is regulated by intricate combinatorial networks (Wilson *et al*, 2010) which maintain self-renewal, lineage specification, differentiation, and apoptotic programs to maintain a steady state with physiologically advantageous populations of blood cells.

Initial characterisation of haematopoiesis by flow cytometry of bulk cell populations employed cell surface markers to define a strictly hierarchical tree structure, starting with Lin^- $Sca\text{-}1^+$ $c\text{-}kit^+$ (LSK) $CD150^+$ $CD48^-$ long-term HSCs (LT-HSCs) at the top of the hierarchy (Kiel *et al*, 2005), differentiating down to LSK $CD34^+$ short-term HSCs (ST-HSCs) (Yang *et al*, 2005), and then making a cell fate decision to bifurcate into either a common lymphoid precursor (CLP) (Lin^- $IL\text{-}7R^+$ $Thy\text{-}1^-$ $Sca\text{-}1^{lo}$ $c\text{-}kit^{lo}$; Kondo *et al*, 1997) or a common myeloid precursor (CMP) (Lin^- IL-

7R$^-$ Sca-1$^-$ c-kit$^+$ FcγR^{lo} CD34$^+$; Akashi *et al*, 2000). Further cell fate choices to differentiate into either megakaryocyte erythroid progenitor (MEP) (CD16/32^{lo} CD41$^-$ CD150$^+$ CD105^{lo} ; Pronk *et al*, 2007) or granulocyte monocyte progenitor (GMP) (CD16/32^{hi} CD41^{lo} ; Akashi *et al*, 2000) cells were followed finally by terminally differentiated cells. This series of discrete steps was redefined as additional surface markers were identified, by the identification of macrophages and dendritic cells arising from early lymphoid progenitors (Doulatov *et al*, 2010), the finding of lymphoid-primed multipotent progenitor (LMPP) cells in the LSK Flt3$^+$ compartment lacking erythro-megakaryocytic potential (Adolfsson *et al*, 2005), and the classification of functionally distinct lineage-primed multipotent progenitors (MPPs) (Pietras *et al*, 2015).

In fact, alongside these bulk population studies, more precise experiments were redefining the architecture of the haematopoietic landscape from as early as 2013, when single cell transplantation studies suggested that lineage restriction can be found in mouse HSCs (Naik *et al*, 2013, Yamamoto *et al*, 2013). John Dick and colleagues confirmed this effect in humans by observing divergence of erythro-megakaryocytic lineage from bone marrow HSC-MPPs (Notta *et al*, 2016). Extensive granddaughter-tracing experiments have found that most cells inherit ancestral lineage bias from early progenitor cells including HSCs, that bias is amplified by bias-inheriting progeny, but also that lineage bias switching can infrequently be observed (Lee *et al*, 2017). This same study suggested that dendritic cell lineage bias can be identified in HSCs, is characterised by specific transcriptional programs correlated with relative IRF8:PU.1 ratios, and is heritable and reinforced over progressive cell division.

It is important when performing *in vitro* experiments to recognise that cell-extrinsic signalling from the bone marrow microenvironment also plays a significant role in HSC differentiation pathways. HSCs home in to distinct endosteal (Calvi *et al*, 2003) and vascular (Kiel *et al*, 2005) niches with varied cell composition such as osteoblasts and mesenchymal stromal cells. This leads to heterogeneous HSC-niche interactions with diversity provided by extra-cellular signals from cytokines, cytokine-receptor profiles of individual HSCs, and cell-to-cell interactions. Cytokine-mediated modulation of HSC cell fate decisions is well characterised, for instance thrombopoietin signalling is essential for megakaryocyte-biased HSCs (Sanjuan-Pla *et al*, 2013). In addition, the biophysical properties of the niche are heterogeneous throughout the bone marrow, and it has been shown that LSK HPC cell fate can be modulated by varying the extracellular collagen matrix stiffness (Chitteti *et al*, 2015).

1.1.2 LMPP cells

Downstream of HSCs, the Jacobsen group identified LSK $Flt3^+$ LMPP cells, challenging the prevailing consensus position that HSCs committed strictly to common myeloid or common lymphoid progenitor lineages (Adolfsson *et al*, 2005). Human LMPP cells are Lin^- $CD34^+$ $CD38^-$ $CD90^{-/lo}$ $CD45RA^+$ $CD10^-$ and have granulocytic, monocytic and lymphoid potential, but lack erythro-megakaryocytic potential. More recently, human LMPP cells have been divided by CD10 and CD45RA cell surface markers into two subpopulations, one with almost entirely lymphoid potential and another one with lympho-myeloid potential (Karamitros *et al*, 2018). In the same study, single cell RNA-seq has suggested heterogeneity in human lympho-myeloid progenitors at the level of the single cell, and that the transcriptional signature of LMPP cells exists in a continuum with granulocyte-macrophage progenitor (GMP) and multi-lymphoid progenitor (MLP) populations (Karamitros *et al*, 2018).

1.1.3 Hoxb8-FL cells

The Hoxb8-FL cell line has been employed in our experiments to model pre-leukaemic changes, and it is therefore useful to place it within the hierarchical model of haematopoiesis and characterise its phenotype and functionality. Hoxb8-FL cells demonstrate phenotypic and functional similarities to LMPP cells, but have a distinct lin^- $Sca\text{-}1^-$ $c\text{-}kit^+$ $cd34^+$ $Flt3^+$ cell surface marker profile, suggesting that they lie downstream of $Sca\text{-}1^+$ LMPP cells (Redecke *et al*, 2013). Principal component analysis (PCA) of microarray-based gene expression clusters Hoxb8-FL cells in between myeloid and lymphoid branches extending down from LMPP cells (Redecke *et al*, 2013). Conditional Hoxb8 expression by fusing the N terminus of Hoxb8 to the estrogen-binding domain of the estrogen receptor and expressing the fused cDNA in lineage-negative progenitor cells as a retroviral expression vector was first published in 2006 as a method to immortalise neutrophil or macrophage progenitors in medium containing stem cell factor (SCF) or granulocyte-macrophage colony-stimulating factor (GM-CSF) (Wang *et al*, 2006). Redecke *et al* cultured these cells in the presence of estrogen and Flt3L, and found that these conditions promoted exponential growth of blast-like Hoxb8-FL cells (Redecke *et al*, 2013). Upon withdrawing estradiol, Hoxb-FL cells differentiated over 6 days into a bi-phenotypic dendritic cell (DC) population of $CD11b^+$ $CD11c^+$ MHC class II^+ $B220^-$ $PDCA1^-$ conventional dendritic cells (cDCs) and $CD11b^-$ $CD11c^+$ MHC class II^+ $B220^+$ $PDCA1^+$ plasmacytoid dendritic cells (pDCs). Further differentiation in Flt3L demonstrated both type 1 and type 2 cDCs.

Removing estradiol and replacing Flt3L with GM-CSF caused Hoxb8-FL cells to differentiate into a mixed population of $CD11b^+$ $CD11c^+$ MHC class II^+ $B220^-$ dendritic cells and $Gr1^{high}$ $CD11c^-$ MHC class II^- granulocytes, whereas replacing Flt3L with macrophage-colony stimulating factor (M-CSF) resulted in adherent cells with a $CD11b^+$ $Gr1^-$ MHC class II^- cell surface marker profile. Hoxb8-FL cells were cultured continuously over 6 weeks and their lineage potential tested by clonal assays and limiting dilution assays, and were found to constitute a homogenous population of cells which maintained multilineage potential stably (Redecke *et al*, 2013).

Furthermore, the multi-lineage cells differentiated from the Hoxb8-FL cell line were functionally validated by the Häcker group. Functional analysis of Hoxb8-FL cells showed that CpG-DNA-mediated stimulation caused pDCs to produce interferon alpha (IFNα) and cDCs to produce IL-12 p40. Type 1 and 2 cDCs triggered proliferation of MHC class II and MHC class I-restricted T cells after incubation with soluble ovalbumin antigen. M-CSF-driven Hoxb8-FL-derived macrophages demonstrated phagocytic properties and nitric oxide production upon provocation with lipopolysaccharide and IFN-γ (Redecke *et al*, 2013).

In vivo transplantation of Hoxb8-FL cells into lethally irradiated mice generated three populations of $CD11b^+$ cells with variable Gr1 expression, $CD11b^-$ $CD11c^-$ $B220^+$ B lymphocytes and $CD3^+$ T lymphocytes in peripheral blood, and also demonstrated myeloid and lymphoid potential *in vivo* in an organ-specific and time-dependent manner. Characterisation of Hoxb8-FL-derived immune cells from spleen and peripheral blood demonstrated that Hoxb8-FL-derived cells *in vivo* corresponded functionally to primary LMPP-derived cells. The LMPP-like phenotype was further confirmed by absent megakaryocyte and erythroid potential after seeding cells in medium which supported development of these cell lines. Loss of immature cells over time suggested that Hoxb8-FL cells do not have self-renewal capacity.

In summary, the Häcker group has developed a remarkably robust and plastic LMPP-like cell line which is conditionally immortalised, recapitulates myeloid and lymphoid phenotype, and is capable of stable integration of mutant gene constructs which can be employed to study leukaemia-associated perturbations of differentiation pathways.

1.1.4 Dendritic cells

Hoxb8-FL cells can differentiate in Flt3L conditions into both conventional and plasmacytoid DCs, however it should be emphasised that the characterisation of DC identity, subtypes and function continues to evolve rapidly. It is therefore useful to summarise current understanding of DCs and DC subsets, DC function, and DC transcriptional signatures, as well as how recent work including scRNA-seq is redefining the classical hierarchy of DC identities.

Dendritic cells can broadly be defined in functional terms as ubiquitous sentinel cells which are capable of sensing early danger signals through a broad array of pattern recognition receptors, and which can bridge the innate and adaptive immune systems by *de novo* processing and presentation of antigen (Guermonprez *et al*, 2002). The classical hierarchical model of haematopoiesis describes MPP commitment to the DC lineage by differentiation into the DC-restricted common DC progenitor (CDP), followed by further DC specification into cDC or pDC subsets (Fogg *et al*, 2006, Onai *et al*, 2007, Onai *et al*, 2010). Type 2 $CD4^+$ $CD11b^+$ $SIRP\alpha^+$ cDCs specialise in phagocytosis and endocytosis of foreign peptides before presenting exogenous antigen via MHC class II pathways to $CD4^+$ helper T cells, initiating antigen-specific antibody production by B cells (Pooley *et al*, 2001). By contrast, type 1 $CD8^+$ $XCR1^+$ cDCs capture endogenous peptide antigen from intracellular cytosolic degradation of proteins and typically present endogenous epitopes on MHC class I molecules to $CD8^+$ cytotoxic T cells. In addition, type 1 cDCs are specialised for antigen cross-presentation, permitting the presentation of exogenous antigens (including tumour), which are normally presented on MHC class II molecules, to activate cytotoxic T cells through MHC class I pathways (Belz *et al*, 2005, Hildner *et al*, 2008, Bachem *et al*, 2012). $IRF8^+$ $B220^+$ $BST2^+$ pDCs are also capable of antigen presentation but are specialised for response to viral infection and produce large quantities of interferon-alpha (IFN-α) in response to stimulation via Toll-like receptors 7 and 9 (Reizis *et al*, 2011).

TF networks participate in differentiation programs associated with DC cell fate decisions. For instance, the ID2/E2-2 axis regulates differentiation towards pDCs and cDCs (Swiecki *et al*, 2015). In addition, TFs play significant roles in DC function, such as RUNX2-mediated egress of pDCs from the bone marrow (Sawai *et al*, 2013, Chopin *et al*, 2016). Type 1 cDCs are regulated by Gata2, PU.1, Gfi1, Irf8 and Batf3 (Onodera *et al*, 2016, Carotta *et al*, 2010, Rathinam *et al*, 2005, Grajales-Reyes *et al*, 2015), whereas type 2 cDCs are regulated by IRF4, NOTCH2 and

KLF4 (Murphy *et al*, 2016, Guilliams *et al*, 2016). Key roles in DC lineage specification are played by PU.1, Irf8 and C/EBPα. Elevated PU.1 levels cause higher order chromatin remodelling to bring a -50 kb *IRF8* enhancer into close contact with the *IRF8* promoter, favouring DC lineage commitment (Schonheit *et al*, 2013). Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBPα and PU.1 respectively demonstrates the developmental plasticity of cell identity (Laiosa *et al*, 2006). More recently, Rosa *et al* have used PU.1, Irf8 and Batf3 TFs to reprogramme mouse and human fibroblasts into DCs which can cross-present antigen to T cells (Rosa *et al*, 2018). The Tenen group has demonstrated that C/EBPα is necessary for differentiation from HSPCs to dendritic cell populations both *in vitro* and *in vivo*, and discovered C/EBPα-dependent genes (eg *ELANE*, *CYBB*, *GATA3*) which are required for early DC differentiation (Welner *et al*, 2013).

The Immunological Genome (ImmGen) consortium has employed large data sets to decipher transcriptional programs of DC lineage specification in mouse cells. Affymetrix microarray expression profiles of cells in the ImmGen database have identified Ikaros and Flt3 as essential TFs for the differentiation of pDCs and cDCs, as well as TFs which regulate DC diversification into type 1 cDCs (BATF3, IRF8), type 2 cDCs (IRF4, NOTCH2) and pDCs (E2-2, SPI-B, IRF8) (Miller *et al*, 2012). In addition, the same study has established a core cDC signature of 24 genes which are differentially expressed in cDCs compared to macrophages, as well as unique gene signatures associated with pDCs and cDCs. Furthermore, Miriam Merad and colleagues from ImmGen found that most of the TF genes which increased expression during commitment of macrophage-dendritic cell progenitors (MDPs) to CDPs but not to monocytes, also encoded TFs such as RUNX2 and KLF8 which are known to control pDC lineage specification, suggesting that CDP differentiation into pDCs constitutes a 'default' pathway (Miller *et al*, 2012). The ImmGen gene expression dataset has also been used derive gene expression modules for DC subsets (Pandey *et al*, 2013).

By contrast with the ImmGen consortium approach, Lin *et al* obtained MPPs, CDPs, cDCs and pDCs by FACS sorting mouse bone marrow cells, and then analysed gene expression and PU.1 binding/histone modification at distinct stages of DC differentiation. This approach revealed that stage-specific changes in gene expression for key DC regulators such as IRF8 and TCF4 are associated with histone modifications in promoter and enhancer regions (Lin *et al*, 2015).

Recent work has challenged traditional hierarchical DC identities. Lee *et al* analysed the quantitative potency of 2,247 HSC and MPP clones derived from $CD34^+$ human cord blood cells, and found no evidence of significant populations correlating with MDPs or CDPs, but rather found lineage priming of DCs marked by *IRF8* expression during early haematopoiesis (Lee *et al*, 2017). The Tamura group have recently employed *IRF8* knock-in and knock-out mice to identify epigenetically-primed DC lineage bias in $IRF8^+$ LMPP cells mediated by altered chromatin accessibility (Kurotaki *et al*, 2019). Of particular interest, FACS sorting of DCs and monocytes from peripheral blood mononuclear cells of healthy donors, followed by single cell RNA-seq of single DCs using the Smart-seq2 protocol, identified a novel Axl^+ $Siglec6^+$ DC subset which was phenotypically and functionally distinct from classical pDCs (Villani *et al*, 2017).

1.2 Malignant haematopoiesis

1.2.1 Pre-leukaemic cell of origin

Leukaemogenesis is a stepwise process which requires sequential mutations and clonal expansion to cause disease. 10% of elderly patients have clonal haematopoiesis of indeterminate potential (CHIP), defined as somatic mutations in the haematopoietic compartment but no evidence of haematological malignancy (Steensma *et al*, 2015). Predictive models have been developed to distinguish cases of age-related clonal haematopoiesis which are more likely to progress to AML, finding that *TP53* mutations are more likely to predict AML than *DNMT3A* or *TET2* mutations in healthy individuals. It has been suggested that a predictive model based on somatic point mutations and variant allele frequencies can predict AML years or even a decade before clinical diagnosis (Abelson *et al*, 2018). In addition, pre-leukaemic mutations which persist in remission are necessary but not sufficient to recapitulate leukaemia, predisposing individuals to relapse (Rothenberg-Thurley *et al*, 2018).

Gene mutations, rearrangements and altered expression profiles are instructive with regard to the stepwise leukaemic evolution of AML. Nerlov and colleagues employed a knock-in mouse model to show that *CEBPA* mutations cause pre-malignant expansion in haematopoietic stem and progenitor cell (HSPC) compartments prior to the development of overt leukaemia (Kirstetter *et al*, 2008). In familial AML, germline mutations in the N-terminal domain of *CEBPA* can be identified in patients, who subsequently sustain somatic *CEBPA*

mutations in the C-terminal domain as second events (Pabst *et al*, 2008). Furthermore, several groups have shown that patients with *CEBPA*-mutated AML carry the same bi-allelic mutations on relapse (Shih *et al*, 2006, Kirstetter *et al*, 2008), suggesting that *CEBPA* mutations represent an early pre-leukaemic event.

MLL gene rearrangements have been identified in pairs of monozygotic twins, suggesting that pre-leukaemic events can exist *in utero* (Ford *et al*, 1993). Similarly, there is evidence for the prenatal origin of the t(8;21) AML1-ETO translocation (Wiemels *et al*, 2002). In addition, the translocation t(8;21) persists in patients with long-term remission (Miyamoto *et al*, 2000), suggesting that this gene rearrangement is a pre-leukaemic event in the sequential acquisition of mutations.

Besides mutations and gene rearrangements, both *GATA1* knockdown (Shimizu *et al*, 2004) and reduced PU.1 expression (Rosenbauer *et al*, 2004) have generated pre-leukaemic phenotypes in mouse models. Similarly, epigenetic modifications such as hypermethylation play a role in pre-leukaemic changes such as HSC expansion, as evidenced by experimental data from *DNMT3A*, *TET2* and *IDH1* mutation models (Challen *et al*, 2011, Moran-Crusio *et al*, 2011, Sasaki *et al*, 2012).

1.2.2 Leukaemia initiating cells

John Dick and colleagues transplanted lineage-negative $CD34^+$ $CD38^-$ AML cells into non-obese diabetic mice with severe combined immunodeficiency disease (NOD/SCID mice) and found that they homed to the bone marrow and proliferated with a leukaemic cell morphology similar to that seen in patients (Lapidot *et al*, 1994). Bonnet and Dick then established a hierarchical model of cancer, where the SCID leukaemia-initiating cells (LIC) in all sub-types of AML were $CD34^+$ $CD38^-$, suggesting that the leukaemic clone originated in normal primitive cells rather than in committed progenitor cells (Bonnet *et al*, 1997).

The emergence of more immunodeficient mouse strains, including NOD SCID Gamma (NSG) mice which lack natural killer (NK) cells, permitted further characterisation of LICs. Taussig *et al* showed that $CD34^-$ $CD38^+$ progenitors sorted from patients with *NPM*-mutated AML were able to recapitulate a leukaemic phenotype in an IVIG-treated NOD/SCID/IL2r$\gamma^{-/-}$ mouse model (Taussig *et al*, 2010). Gradually a body of work has established that LICs are

heterogeneous, for instance the Weissman group retrovirally transduced sorted HSC, CMP and GMP populations with MSCV/MLL-ENL-IRES-GFP and transplanted them into irradiated mice, showing that myeloid progenitors as well as HSCs were capable of generating a leukaemic phenotype (Cozzio *et al*, 2003). Krivtsov *et al* found that introduction of the MLL-AF9 fusion gene into GMPs resulted in LICs with a global gene expression profile and immunophenotype very similar to normal GMPs but with a leukaemia self-renewal signature (Krivtsov *et al*, 2006). Similarly, GMPs were identified as candidate LICs in blast-phase chronic myeloid leukaemia (CML) (Jamieson *et al*, 2004).

More recently, Paresh Vyas and colleagues have shown that ~80% of leukaemic stem cells from $CD34^+$ AML demonstrate the immunophenotype and gene expression profile of LMPP-like or GMP-like progenitors, rather than of HSCs, suggesting that most leukaemic stem cells in primary $CD34^+$ AML are progenitor cells which acquire aberrant self-renewal capacity (Goardon *et al*, 2011). We therefore hypothesise that LMPP-like Hoxb8-FL progenitor cells provide a stable and reproducible cellular model that can be used to study gene expression changes implicated in leukaemia initiation.

1.2.3 AML as a clinical entity

Acute myeloid leukaemia (AML) is the most common adult leukaemia with an incidence of 15 per 100,000 at age 75 and a median age at diagnosis of 67 years (Roug *et al*, 2014). Our understanding of leukaemia has evolved significantly since Alfred Donné's speculation in 1844 that the "mucous globules" he observed microscopically in a patient with splenomegaly were due to differentiation arrest of white blood cells (Kampen, 2012). Since those early descriptions, apart from progress in clinical and morphological diagnosis, major advances have unfolded from the 1970s onwards in diagnostics, prognostics, and therapeutic options. Today we understand that AML results from the stepwise acquisition of somatic mutations (Figure 1.1) which lead to dysregulation of cell cycle control, limitless replicative potential, genomic instability, altered micro-environment, and ultimately the clonal expansion of immature myeloid cells in differentiation arrest.

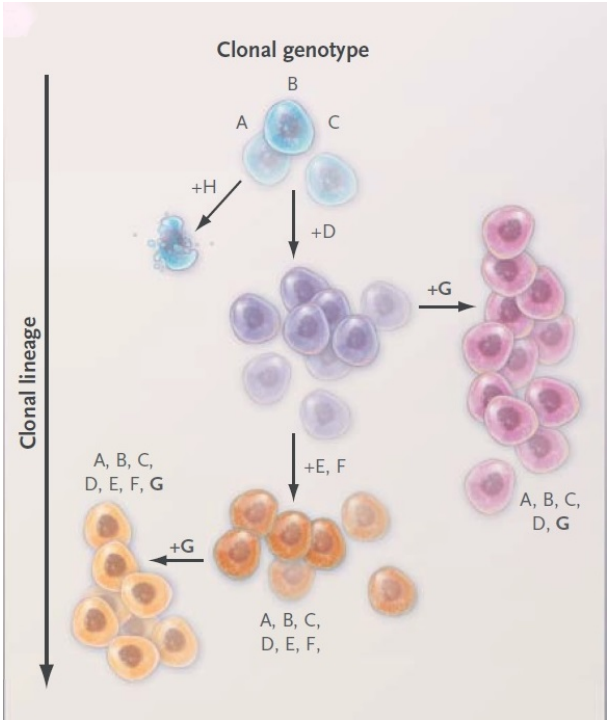


Figure 1.1 Clonal relationships (adapted from Aparicio *et al*, 2013)

AML represents a haematological emergency and morphology can provide an immediate diagnosis in the acute clinical setting, with the World Health Organization (WHO) definition in most cases requiring 20% myeloid blast cells in peripheral blood or bone marrow (Vardiman *et al*, 2002). Initial classification of AML continues to employ the French-American-British (FAB) morphological criteria established in 1976 (Bennett *et al*, 1976). The identification and quantification of myeloid blasts is now commonly confirmed by flow cytometry to identify human AML-associated immunophenotypes including CD13, CD33 and CD117 (Bene *et al*, 1995). Cytogenetic abnormalities such as inv(16) and t(15;17) provide a diagnosis of AML even in the absence of 20% myeloid blasts on morphological examination.

More recently, molecular mutations have been assimilated into diagnostic criteria for AML (Roug *et al*, 2014), with *NPM1* and *CEBPA* mutations designated as AML entities in the 2008 WHO classification and the addition of *RUNX1* mutations in its 2016 revision (Vardiman *et al*, 2009, Arber *et al*, 2016). In addition, whole genome sequencing (WGS) of patients with AML has uncovered mutations affecting epigenetic machinery, including *DNMT3A* (Ley *et al*, 2010) and *IDH1* (Mardis *et al*, 2009). WGS of 200 patients with AML identified approximately 23 driver mutations (Cancer Genome Atlas Research Network, 2013), a relatively low number

which suggested that AML may comprise a simpler mutational model than solid tumours (Tenen, 2003). More recently, Peter Campbell and colleagues described 5234 driver mutations across 76 genes or genomic regions, with two or more driver mutations identified in 86% of patients with AML (Papaemmanuil *et al*, 2016). The challenge of the post-genomic age is to elucidate elegantly how these gene perturbations interact with each other within transcriptional networks to generate leukaemic phenotypes.

Prognostic information guides clinicians in risk-adapted therapeutic decisions after induction of remission, with poor-risk patients who are fit for aggressive therapy proceeding to stem cell transplantation. By contrast, favourable-risk patients normally proceed to consolidation chemotherapy after successful first remission. Cytogenetics remains the most important disease-specific risk factor for AML, and poor-risk karyotypes include t(9;22), monosomy 5 and complex karyotype (Grimwade *et al*, 1998). However, approximately 60% of patients have intermediate risk cytogenetics, and in this cohort of patients molecular mutations provide important prognostic information: *NPM1* and *CEBPA* with double mutation (*CEBPA*^{dm}) carry favourable prognosis, whereas *FLT3-ITD* carries poor prognosis (Papaemmanuil *et al*, 2016). Wakita *et al* have described the concept of Complex Molecular Genetic Abnormalities (CMGAs) in AML (Wakita *et al*, 2016), arguing that the poor prognosis of *FLT3-ITD* is not an intrinsic effect of the mutation itself but rather is because *FLT3-ITD* commonly co-exists in CMGAs with multiple concurrent mutations, whereas the converse is true of *CEBPA*^{dm} mutations.

Therapeutics pose a significant challenge in the management of AML. Initial induction of remission continues to be by a combination of antimetabolite and anthracycline chemotherapy which was first proven to have efficacy in 1973 (Yates *et al*, 1973). Relapse remains a significant clinical problem with current therapies, due either to the founding clone expanding through therapy or else the development of additional mutations in the founding clone (Ding *et al*, 2012). Unfortunately, progress in diagnostics and prognostication has not been matched by the emergence of effective targeted molecular therapies for AML. In this regard, WGS has uncovered the significant driver mutations which initiate AML, but fails to address the complex transcriptional dysregulation by which initial mutational simplicity brings about a highly heterogeneous and complex phenotype. In consequence, overall survival for

patients with AML at 5 years remains low for younger patients and dismal for patients over the age of 60 who constitute the majority of patients with AML (Dickson *et al*, 2016).

Molecular therapies such as the FLT3 inhibitor Quizartinib are currently undergoing development in phase III clinical trials (Cortes *et al*, 2019). Pharmacological inhibition of transcription has traditionally been problematic, but the emergence of therapies such as BET inhibitors (Zuber *et al*, 2011) suggests that transcriptional dissection can lead to novel therapeutics as well as an increased understanding of the pathogenesis of AML. The effectiveness of immunotherapy in leukaemia has thus far been limited. Allogeneic stem cell transplants provide evidence that T lymphocytes can eradicate leukaemic cells, and donor lymphocyte transfusions were able to treat relapsed CML before the development of tyrosine kinase inhibitors (Kolb, 1998). More recently CAR T-cell therapy has been licenced for use in refractory B-ALL (Rose, 2017), comprising an antibody-like domain targeting a specific antigen such as CD19, a transmembrane domain, and an intracellular signalling domain which activates T-cells. Early excitement around DC vaccination was generated by the finding that a single injection of antigen-pulsed DCs in healthy subjects was able to induce a T-cell response (Dhodapkar *et al*, 1999), and myeloid malignancies were seen as a prime target (Schurch *et al*, 2013, Tesfatsion *et al*, 2014, Pyzer *et al*, 2014). Broadly speaking, DC vaccine therapy comprises of isolating autologous or allogeneic DCs, incubating with peptides or tumour cells to load MHC molecules on the surface of the DCs with antigen, and then delivering DCs to the lymph nodes to engage a T cell response. Clinical trials of DC therapy in the setting of AML have mostly focused on adjuvant therapy post-allogeneic transplant (Wang *et al*, 2018), on treatment of refractory/relapsed AML where conventional therapy has failed (Li *et al*, 2006), or on delaying relapse for patients in remission (Khoury *et al*, 2017). Overall, these studies have demonstrated limited clinical response in patients with AML, and work is ongoing to develop DC-based therapies with increased immunostimulatory activity.

1.3 Transcription factor regulation of haematopoiesis

Despite the fact that relatively few recurrently mutated genes have been identified in patients with *de novo* AML (Cancer Genome Atlas Research Network, 2013; Papaemmanuil *et al*, 2016), leukaemic phenotypes are enormously diverse. To some extent, this can be explained by complexity at the level of transcription, caused both by the stochastic nature of transcriptional

burst kinetics (Raser *et al*, 2004, Suter *et al*, 2011) and by malignant perturbation of intricate gene expression networks which regulate cell state stability of haematopoietic stem cells and progenitors during normal haematopoiesis (Schutte *et al*, 2016).

The mechanism for transcriptional regulation derives from the interaction between transcription factors (TFs) and *cis*-regulatory elements on DNA, whereby TFs program cellular phenotypes. Indeed, it has been established that a single TF can reprogram cells, for instance converting fibroblasts to contracting myocytes (Davis *et al*, 1987). However, increasingly, the attention of the scientific community has focused on the transcriptional paradigm as a network of TF combinations which are spacio-temporally interconnected and which often act as TF complexes rather than in isolation (Wilkinson *et al*, 2017). Riddell *et al*, for instance, used six TFs to reprogram mature pro-B cells into HSC-like cells in a mouse model (Riddell *et al*, 2014).

Two distinct, but mutually complementary, approaches have been developed to analyse the topology of Gene Regulatory Networks (GRNs). Firstly, a 'bottom-up' approach typically establishes a single TF-DNA interaction by functional studies, and then either builds outwards from this central hub or else places it as a new node within an existing network paradigm. An example of this process was the Stem cell leukaemia (Scl) node, required for HSC specification, being placed within a highly recursive gene regulatory circuit involving GATA2, FLI1 and the enhancers *SCL* +19 and *FLI1* +12 (Pimanda *et al*, 2007). Secondly, technology such as Chromatin ImmunoPrecipitation with massive parallel DNA Sequencing (ChIP-seq) combined with computational approaches has enabled a 'top-down' approach where genome-wide binding maps suggest novel TF interactions which can then be confirmed by functional studies. This process led, for instance, to the identification of regions mapping to sixteen TF genes as targets of the basic helix-loop-helix TF SCL in Haematopoietic precursor cell-7 (HPC-7) cells, and then functional *in vivo* validation of these putative regulatory elements by inserting them into *lacZ* reporter constructs in a transgenic mouse model (Wilson *et al*, 2009). More recently, a 'top-down' approach has permitted analysis of combinatorial transcriptional control by ten major TFs in blood stem/progenitor cells, including a novel protein-protein interaction between GATA2 and RUNX1 which was functionally validated by demonstrating the non-viability of GATA2/RUNX1 compound heterozygous mice (Wilson *et al*, 2010). Increasingly, next generation technology and computer modelling will take a central role in representing

patterns of behaviour within complex TF networks [Figure 1.2], but these models continue to require functional validation by *in vitro* and *in vivo* experiments.

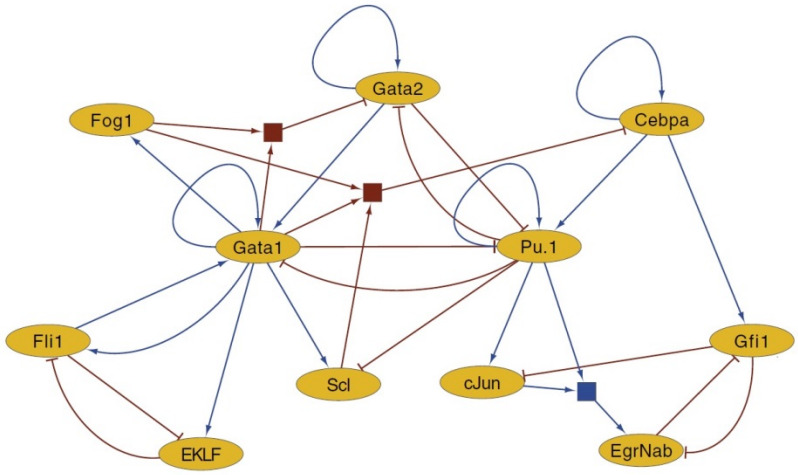


Figure 1.2 Boolean model of core TF networks active in common myeloid progenitors (adapted from Moignard *et al*, 2013)

Perhaps as challenging as the vast possibilities of transcriptional network interactions presented by approximately 25,000 protein-coding genes in the murine or human genome, is the intriguing question of the relatively small number of cell fates which result from these interactions. This can be partially explained by network motifs which control information flow within GRNs and which can be inferred by perturbation studies. Notwithstanding, ChIP-seq clearly describes an extraordinary number of binding interactions between TF and DNA. The observation that binding patterns for a given TF can differ widely with cellular context led to the proposition that the majority of binding events are 'opportunistic' and only those patterns which are constant between cellular contexts are 'functional' (Wei *et al*, 2011). However, work on transcriptional interactions in two distinct cell types has suggested that cell type-specific binding regions, far from being redundant, demonstrate specific patterns of consensus sequence motifs which may participate in DNA accessibility to shared TFs within a given cellular context (Calero-Nieto *et al*, 2014). The emerging picture is of a complex interplay of TF networks operating at more than one level which permits intrinsic master regulator TFs to give consistent cell fates in an extrinsic cellular context. It has further been observed that ChIP-seq presents an excess of putative target genes when data is analysed for single TFs but that biological functionality is more easily established when statistical analysis is employed to identify binding patterns between DNA and TF complexes (Ottersbach *et al*, 2010). Intuitively, it can be hypothesised that if the vast majority of binding events are not redundant, then TFs

must produce limited cell fates by acting within TF complexes which operate in defined network motifs and which have mechanisms to deal with varied cellular context.

Single cell studies have given additional useful information regarding the interactome. Gottgens and colleagues have demonstrated that TF expression is heterogeneous within LT-HSC, LMPP, MEP, GMP and CLP populations, with expression levels being more constant in specific TFs (for instance, RUNX1) than in others (Moignard *et al*, 2013). This suggests a limited tolerance for variation within TF networks. Furthermore, Index Sorting permits mapping of cells back to their original sort location and retrospective interrogation of heterogeneous flow-sorted populations (Schulte *et al*, 2015). Thirdly, single cell studies can go further than positively correlating the expression of two genes, and can suggest whether two genes share a common regulatory mechanism or whether one gene upregulates expression of the other (Moignard *et al*, 2014). For instance, genes X, Y and Z may be upregulated in a cell population, but single cell analysis may indicate that genes X and Z are upregulated in the same single cells, and gene Y is upregulated in distinct single cells within that population, suggesting that genes X and Z are more likely to have a shared regulatory mechanism than gene Y (Bengtsson *et al*, 2005).

Ideally an increased appreciation for the topology of GRNs will encourage and direct the process of developing targeted therapies. For instance, TF network analysis has established MYC as a central hub in the B cell transcriptome (Basso *et al*, 2005), and potentially a promising target for therapy (Soucek *et al*, 2008). In any case, it is clear that the phenotypic output of a given TF, and the consequences of its mutation, can only be understood in the context of its network interactions with other TFs and with *cis*-regulatory elements on DNA.

1.4 CCAAT/enhancer binding protein (C/EBP)α

1.4.1 C/EBPα overview

C/EBPα is the product of an intronless gene located on chromosome 19q in humans and on chromosome 7 in mice. It functions as a bZIP TF and originally was found to regulate hepatocyte and adipocytes genes in rat liver (Cao *et al*, 1991). Pabst *et al* first identified dominant negative *CEBPA* mutations in human AML (Pabst *et al*, 2001) and intense research followed the finding that C/EBPα plays a key role in both myeloid differentiation and cell cycle

arrest (Timchenko *et al*, 1996, Wang *et al*, 2001). Interestingly, although *CEBPA* is described as a tumour suppressor gene (Reckzeh *et al*, 2010), *CEBPA* expression at the GMP stage [Figure 1.3] is required for myeloid lineage commitment and hence for the generation of AML (Ohlsson *et al*, 2016). Furthermore, evidence suggests that C/EBPα may act as a pioneer factor, opening up chromatin and increasing accessibility for other TFs (Ohlsson *et al*, 2016).

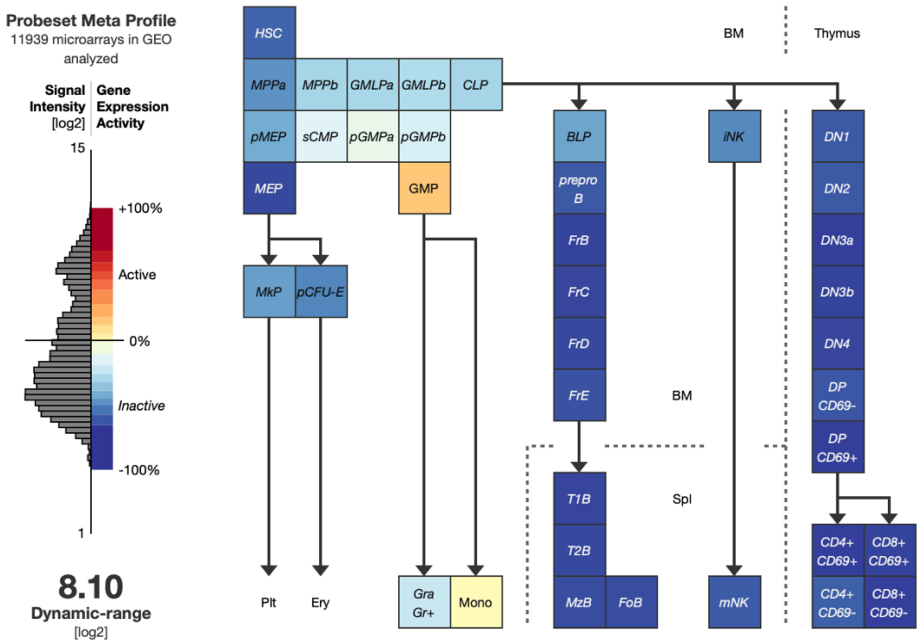


Figure 1.3 Gene expression activity of *CEBPA* in 39 distinct highly purified murine haematopoietic cells, based on varied microarray data as a common reference (from https://gexc.riken.jp/models/3/genes/Cebpa ; Seita *et al*, 2012)

1.4.2 C/EBPα structure and function

CEBPA produces a full 42-kDa C/EBPα isoform (henceforth referred to as p42) which contains two transactivation domains (TADs) to activate transcription at the N-terminus, as well as basic region (BR) and leucine zipper (LZ) domains, for DNA-binding and dimerisation respectively, at the C-terminus [Figure 1.4]. p42 is the primary physiological isoform, but a distinct 30-kDa isoform (p30) can also be translated from an alternative ATG start codon within *CEBPA*. p30 has intact BR-LZ DNA-binding function, but reduced transactivation properties. Specifically, p30 omits the first 117 amino acids of p42, and therefore lacks TAD1 which in p42 interacts with the TATA-Binding Protein and hence with the RNA polymerase II transcription holocomplex. In addition, p30 does not exert cell cycle control (Ohlsson *et al*,

2016) and p30 heterodimers have reduced DNA binding *vis-à-vis* p42 homodimers (Cleaves *et al*, 2004).

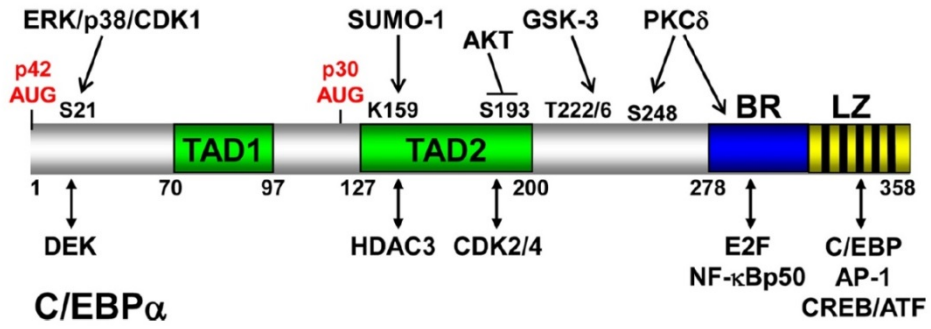


Figure 1.4 Structure of C/EBPα showing transactivation domains, basic region, leucine zipper, initiating AUG residues, protein modifications and protein interactions (from Friedman, 2015)

Dimerisation is an intrinsic feature of the C-terminal basic zipper (bZIP) structural motif found in C/EBPα. Hydrophobic leucine residues in an α-helical structure interdigitate with parallel residues in a partner protein at their respective LZ domains, and the two dimerised proteins then bifurcate at their BR domains to bind the major groove of DNA. The standard model for C/EBPα is the p42 homodimer, but functional diversity is provided by an alternative p30 homodimer and by heterodimerisation with other LZ transcription factors including IRF8 (Kurotaki *et al*, 2014) and AP-1 (Cai *et al*, 2008). Co-immunoprecipitation studies suggest that C/EBPα favours homodimerisation or heterodimerisation with C/EBPβ over JUN or FOS TFs (Cai *et al*, 2008). Noncovalent interactions push a protein towards homodimerisation or heterodimerisation, and a single amino acid variation in LZ structures can select for a particular dimerisation profile (Pogenberg *et al*, 2014). Each configuration binds specific DNA sequence motifs and hence generates a distinct expression profile.

The first major function of C/EBPα is to instruct cell fate within transcriptional regulatory circuits. For instance, both a conditional knock-out mouse model of *CEBPA* (Hasemann *et al*, 2014) and deletion of the +37 kb *CEBPA* enhancer (Avellino *et al*, 2016, Guo *et al*, 2016) resulted in loss of LT-HSCs. C/EBPα plays a critical role in the differentiation pathway from Common Myeloid Progenitors (CMPs) to Granulocyte-Monocyte Progenitors (GMPs) (Nimmo *et al*, 2015). Correspondingly, *CEBPA* -/- mice are deficient in granulocytes and monocytes (Zhang *et al*, 1997, Heath *et al*, 2004). The role of C/EBPα in myeloid gene expression programs

is evidenced by C/EBPα-mediated reprogramming of fibroblasts into myeloid progenitors (Feng *et al*, 2008). The Tenen group have shown that DC progenitor cell fate is also C/EBPα-dependent in a knock-out mouse model (Welner *et al*, 2013).

The second major function is cell cycle control. CEBPs as a TF family act as tumour suppressors in a wide range of malignancies, including AML, lung cancer, skin cancer, hepatocellular carcinoma and breast cancer (Nerlov, 2007). C/EBPα represses the E2F cell cycle regulatory complex, thereby controlling cellular proliferation but also inducing terminal differentiation (Porse *et al*, 2001). C/EBPα may also inhibit the cell cycle by interfering with Cyclin Dependent Kinases CDK2 and CDK4 (Wang *et al*, 2001), although deletion of the CDK2/CDK4 binding domains of *CEBPA* in mice did not alter cell proliferation (Nerlov, 2004).

1.4.3 *CEBPA* mutations

AML has been characterised as a collaboration of class I 'signalling mutations' which activate signal transduction pathways and confer a proliferative advantage on haematopoietic cells, with class II mutations which affect transcription or nuclear function and impair differentiation (Gilliland, 2002). *CEBPA* class II mutations are classified as N-terminal or C-terminal mutations, and each have general properties which can be related to biological structures coded by the *CEBPA* gene [Figure 1.5].

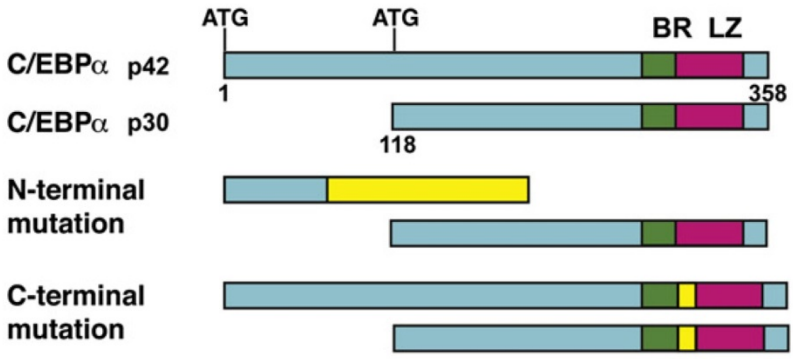


Figure 1.5 Schematic representation of C/EBPα isoforms occurring naturally and as a result of AML-associated mutations (from Bereshchenko *et al*, 2009). N-terminal mutations are frame-shift mutations that block the translation of the 42 kDa isoform (p42) while allowing the 30 kDa isoform (p30) to be expressed. C-terminal mutations generate in-frame insertions/deletions within the basic region-leucine zipper DNA binding domain.

N-terminal frame-shift mutations preserve p30 but prematurely truncate the p42 isoform. Increased p30 expression then inhibits wildtype p42 in a dominant negative fashion (Pabst *et al*, 2001). The overall effect is reduced transactivation function. In fact, most N-terminal mutations lie downsteam of the ATG which initiates synthesis of p42 and therefore lead to expression of p30. In addition, bi-allelic mutations preferentially form p30 homodimers (Ohlsson *et al*, 2016), and therefore p30 is the most common isoform to result from mutated *CEBPA*. To investigate the phenotypic effects of p30 dominance, specific ablation of p42 translation was carried out in a Cre-Lox knock-in mouse model. This resulted in AML with death between 9 and 14 months (Kirstetter *et al*, 2008). Interestingly, GMPs were seen in this disease model, indicating that p30 alone is sufficient for the CMP to GMP transition.

C-terminal in-frame mutations have varied effects on transactivation, but are characterised primarily by altered DNA-binding by the bZIP region of C/EBPα. The C-terminal N321D mutation is located at the 321st codon which lies within the leucine zipper (LZ) domain. LZs are characterised by hydrophobic leucine residues at every *d* position in a $(abcdefg)_n$ amino acid repeat, and they are configured in an α-helical structure which interdigitates with parallel residues from a partner protein at their respective LZ domains. Asparagine residues can specify dimer formation (Zeng *et al*, 1997), and in addition dimerisation profiles are particularly affected by amino acids at positions *a*, *e* and *g* in the LZ heptad repeat (Schuermann *et al*, 1991). *CEBPA* N321D causes asparagine to be replaced by aspartate at position *a* in the heptad repeat, and therefore it would be reasonable to hypothesise that this causes LZ reconfiguration and altered DNA-binding with associated transcriptional effects.

In a seminal study by Bereshchenko *et al*, after generating knock-in 'L' (Lp30) and 'K' (K313 duplication) alleles to achieve competitive transplantation of fetal hepatic cells with both N and C-terminal mutations into lethally irradiated mice [Figure 1.6], homozygous C-terminal mutations were found to skew HSC differentiation towards erythroid lineage, with absence of GMPs (Bereshchenko *et al*, 2009). Nevertheless, LSK cells were increased in comparison to homozygous N-terminal mutations, suggesting that C-terminal mutations expand the HSC pool (Bereshchenko *et al*, 2009). Of note, this same study suggested a syngenic mechanism for bi-allelic N-terminal and C-terminal mutations, comprising the most common *CEBPA* mutation profile in AML: the allele combination causes HSC expansion while retaining myeloid programming.

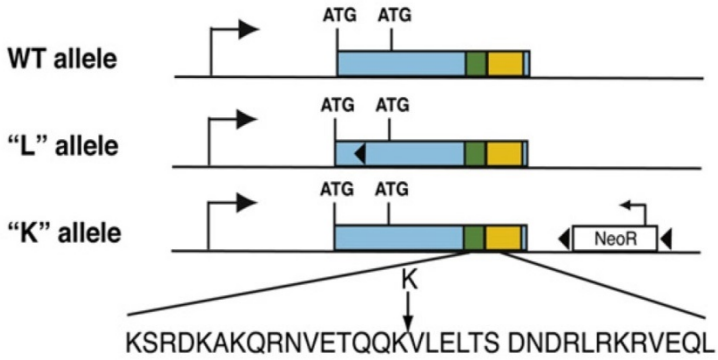


Figure 1.6 Schematic representation of mutant *CEBPA* alleles Closed triangles, LoxP sites; open box, neomycin Resistance cassette; duplicated lysine residue seen in K313KK C-terminal mutation (from Bereshchenko *et al*, 2009)

In a study by Togami *et al*, 32Dcl3 cells were retrovirally transduced with three different C-terminal *CEBPA* mutations and a mock control, then exposed to IL-3 and six days of granulocyte-colony stimulating factor (G-CSF) treatment (Togami *et al*, 2015). The control cells differentiated into neutrophils but the three mutants remained immature, with the K313 mutation inhibiting differentiation most strongly. Subsequently, sublethally irradiated mice were transplanted with mononuclear cells infected with retroviruses harbouring the same mutations. Transplantation with C-terminal mutated cells resulted in leukaemia in all mice, with the shortest median survival time for the N321D mutation [Figure 1.7].

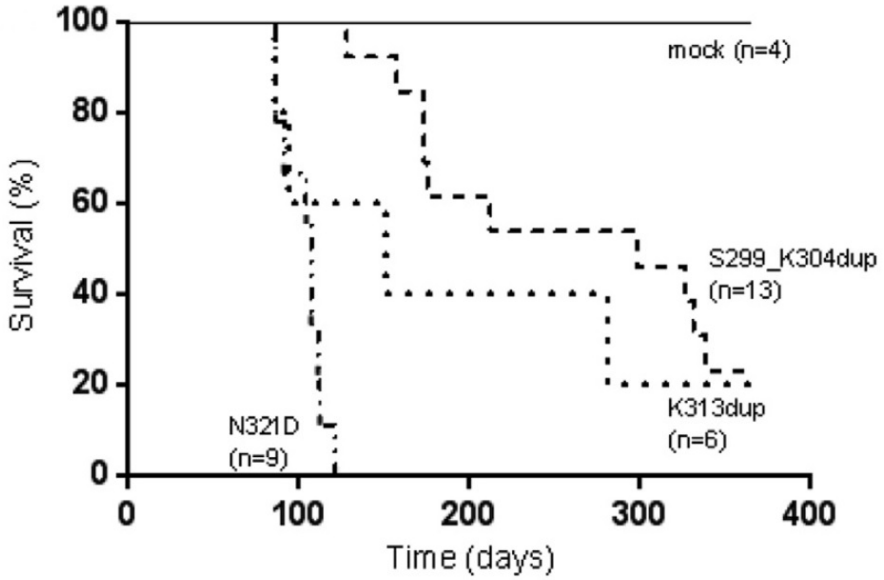


Figure 1.7 Kaplan-Meier analysis for the survival of mice receiving transplants of BM cells transduced with C/EBPα-299_304dup, (299_304dup; n=13), C/EBPα-K313dup (K313dup; n=6), C/EBPα-N321D (N321D; n=9) or pMYs-IG (mock; n=4) (from Togami *et al*, 2015)

The fact that K313 inhibited differentiation most powerfully but the N321D mutation induced a more aggressive leukaemia suggests that C-terminal mutations can cause proliferation as well as inhibit differentiation.

Bereshchenko *et al*'s report that C-terminal mutations in a knock-in mouse model caused expansion of pre-malignant HSCs but required cooperating N-terminal mutations to promote myeloid differentiation into a leukaemic phenotype (Bereshchenko *et al*, 2009) appears to contradict findings by Togami *et al* that retroviral infection of BM mononuclear cells with C-terminal mutations followed by transplantation is able to recapitulate AML *in vivo* without cooperating N-terminal mutations (Togami *et al*, 2015). In another study by Kato *et al*, serial colony assays demonstrated that cells infected with a C-terminal mutation are able to form colonies after 6 rounds of replating and produce a leukaemic phenotype within 4 to 12 months of competitive transplantation into irradiated mice (Kato *et al*, 2011). These divergent results may be reconciled to some extent by hypothesising that the latency of the leukaemic effect found by Kato *et al* suggests additional cooperating mutations. However, the leukaemic phenotype induced by the *CEBPA* N321D C-terminal mutation in work by Togami *et al* had a relatively short latency, and it may be that the divergence in reported results is impacted by different experimental models. Knock-in mouse models express the *CEBPA* mutation via an endogenous promoter, in contrast to retroviral models which transduce exogenous promoters into cells before transplantation. Secondly, knock-in models replace the wildtype (WT) allele with a knock-in mutant allele, rather than integrate a gene mutation into a genome with two WT alleles. Thirdly, semi-random retroviral integration can cause unexpected cellular effects including insertional mutagenesis (Hacein-Bey-Abina *et al*, 2003, Braun *et al*, 2014, Stein *et al*, 2010). However, transplantation models arguably model disease pathology more accurately because only some cells express the mutant whereas knock-in models express the mutation in all haematopoietic cells. In addition, it is also important to remember that *CEBPA* mutant effects can be cell context-specific, for instance a C-terminal mutant has been reported to suppress G-CSF expression and granulocytic differentiation specifically in haematopoietic 32Dcl3 cells and not in 293T cells (Kato *et al*, 2011). Interestingly, the Bonnet group found that intrinsic differences between human and mouse progenitor cells resulted in distinct self-renewal phenotypes on lentiviral transduction of C-terminal mutated *CEBPA* (Quintana-Bustamente *et al*, 2012).

In addition to general characterisation of N- and C-terminal *CEBPA* mutation profiles, two important observations have been made by mouse models. Firstly, *CEBPA* $^{-/-}$ mice lack mature granulocytes but they do not develop AML, suggesting that some degree of *CEBPA* expression is required for the myeloid leukaemic process (Ohlsson *et al*, 2016). Secondly, transplantation of HSC and progenitor cell fractions with *CEBPA* mutations has shown leukaemic engraftment exclusively in committed myeloid cells (Kirstetter *et al*, 2008), suggesting that the leukaemia-initiating cells in *CEBPA*-mutated AML are found downstream of HSCs.

1.4.4 Alternative mechanisms of C/EBPα-mediated AML

Gene silencing through hypermethylation of the *CEBPA* promoter can cause a similar expression profile to *CEBPA* mutations. Interestingly methylation rather than mutation is found commonly in Myelodysplastic Syndromes (MDS) which often presage AML (Wen *et al*, 2015). Post-translational modifications are seen, such as ERK1/2-mediated phosphorylation at serine 21 which changes the conformation of transactivation domains in C/EBPα (Ross *et al*, 2004), thereby altering lineage commitment. Lastly, the interaction between C/EBPα and *CXCR4* is relevant to AML phenotype, with wildtype p42 binding to the *CXCR4* promoter region (Kuo *et al*, 2014). Mutated p30 does not bind to this promoter on ChIP and may downregulate *CXCR4*, increasing AML susceptibility to chemotherapy (Kuo *et al*, 2014).

1.4.5 Combinatorial interactions with other TFs

C/EBPα promotes myeloid transcriptional programs by binding to regulatory elements (promoters or enhancers) for myeloid genes including *KLF5* and *CSF3R* (Federzoni *et al*, 2014), and reintroduction of C/EBPα in leukaemic $CD34^{+}$ stem cells can partially restore myelopoiesis (Schepers *et al*, 2007). It has been noted that *CEBPA* mutations upregulate genes of erythroid differentiation and downregulate genes of myeloid differentiation and homeobox genes (Pabst *et al*, 2009). However, more interestingly for computational analysis, C/EBPα interacts with other TFs in relatively small core circuits which can be hypothesised to participate in larger more complex GRNs (Pimanda *et al*, 2010).

C/EBPα regulates alternative monocyte or neutrophil cell fate in partnership with IRF8. Specifically, IRF8 prevents C/EBPα from binding to its target sequence in monocyte-dendritic cell progenitors, impairing transcription and thereby preventing neutrophil differentiation (Kurotaki *et al*, 2014). Transfection with selected mutations of IRF8 followed by chromatin

immunoprecipitation suggests that IRF8 inhibits C/EBPα activity by interacting with its DNA-binding domain (Kurotaki *et al*, 2014). In line with these results, IRF8 $^{-/-}$ mice demonstrate absent monopoiesis and aberrant over-production of neutrophils (Kurotaki *et al*, 2014). By contrast, IRF8 associates with PU.1 to promote monocyte differentiation (Kurotaki *et al*, 2013). In turn, C/EBPα enhances PU.1 expression by directly binding to its promoter and -14 kb enhancer, and represses PU.1 activity by displacing its co-factor JUN (Reddy *et al*, 2002). C/EBPα can dimerise via its LZ domain with AP-1 proteins and thereby interact with alternative DNA elements which promote monocyte differentiation. In fact, C/EBPα heterodimers with JUN induce monocyte lineage commitment more efficiently than C/EBPα or JUN homodimers (Cai *et al*, 2008). In this regard, one mechanism of increased efficiency may be that the C/EBPα:JUN heterodimer binds and activates PU.1 (Cai *et al*, 2008). Thus IRF8, PU.1, JUN and C/EBPα interact in recursive core circuits to promote monocyte or neutrophil lineage specification and/or differentiation.

Other network motifs remain to be elucidated. For instance, *CEBPA^{dm}* AML has high levels of mutant *WT1* and *GATA2* (Fasan *et al*, 2014), suggesting a possible transcriptional circuit. More recently a novel activation pathway mediated by p53 and KLF4 has been identified with deregulation of the p53-KLF4-C/EBPα axis in AML (Seipel *et al*, 2016).

1.4.6 C/EBPα as a pioneer factor

Evidence is building that C/EBPα may have the ability to interact with closed chromatin, remodelling chromatin and enabling access by other TFs at key regulatory elements. Transient pulsed C/EBPα expression prior to co-expression by TFs OCT4, SOX2, KLF4 and MYC enhances reprogramming of B cells into induced pluripotent stem cells, and C/EBPα pulses after expression of these factors is ineffective (Di Stefano *et al*, 2014). Hasemann *et al* found that conditional acute deletion of *CEBPA* results in histone modifications of genes involved in stem cell function (Hasemann *et al*, 2014). More precise evidence of C/EBPα-mediated chromatin accessibility has recently been provided by Thomas Graf and colleagues who found that pulsed *CEBPA* expression in pluripotent cells generates a cluster of *de novo* ATAC peaks on ATAC-seq, which enriches for the *KLF4* motif, and that interplay between C/EBPα and KLF4 induces chromatin accessibility to the Yamanaka pluripotency factors (Di Stefano *et al*, 2016). These studies suggest that C/EBPα may act as a 'pioneer factor', opening up chromatin for TFs to access DNA and priming myeloid transcriptional programs (Sardina *et al*, 2016).

1.4.7 Clinical features of *CEBPA*-mutated AML

Incidence and prevalence studies need to be interpreted carefully, as recent evidence suggests that incidence can alter significantly both with age (Libura *et al*, 2015) and with ethnicity (Gou *et al*, 2016). Nevertheless, in general, single and double mutations of *CEBPA* account for approximately 7 to 15% of *de novo* cytogenetically normal AML (Preudhomme *et al*, 2002). Of these, about 30% of cases have single allele *CEBPA* mutations, and 70% have bi-allelic mutations most commonly with a N-terminal mutation in one allele and a C-terminal mutation in the second allele. There is consensus that *CEBPA* ^{dm} gives a favourable prognosis in AML (Pabst *et al*, 2009, Wouters *et al*, 2009), recognised in the WHO disease classification (Arber *et al*, 2016). In clinical practice, these patients are therefore more likely to be treated with chemotherapy in the first instance rather than allogeneic stem cell transplantation. By contrast, single-mutated *CEBPA* (*CEBPA* ^{sm}) carries no prognostic significance.

1.5 The Single Cell: A New Paradigm

Cell fate decisions are made by single cells, including rare stem and progenitor cells, and therefore require single cell analysis for precise characterisation. In addition, bulk analysis cannot map cancer heterogeneity, which is one of the principal challenges for developing effective targeted cancer treatments. AML provides an ideal conceptual model for deciphering heterogeneity during clonal evolution because it is characterised by relatively few mutations, particularly at pre-leukaemic phase. Interestingly, heterogeneity also underlies the fact that dendritic cells remain poorly characterised and understood despite playing pivotal roles within the immune system.

Single cell RNA-seq isolates individual cells from a bulk population, lyses those cells to acquire RNA, reverse transcribes RNA into cDNA, and then amplifies the cDNA to generate genome-wide sequencing libraries. Due to the reduced quantity of starting material in each individual cell, the major QC challenge of this technique is technical bias, including accuracy and sensitivity. Cost, numbers of cells, and sequencing depth are other questions which need to be considered in experimental design (Baron-Gale *et al*, 2018). A variety of scRNA-seq methods and protocols have developed to meet these challenges, and the comparative strengths and weaknesses of each method (Ziegenhain *et al*, 2017) need to be considered with regard to the hypothesis being tested. For instance, the 10x Chromium system is automated

and therefore is less prone to manual handling bias, it employs unique molecular identifiers (UMIs), and it applies a 3'-tag or 5'-tag sequencing method. By contrast, Smart-seq2 often captures more genes per cell, it uses spike in controls for RNA normalisation, and can capture full-length RNA.

In addition to the range of available scRNA-seq protocols, several combinatorial technologies have emerged to increase the versatility and power of single cell analytical pipelines. For instance, combining scRNA-seq with mass cytometry or Cytometry by Time Of Flight (CyTOF) offers significant advantages over flow cytometry, including the employment of non-fluorescent metal isotopes to conjugate antibodies, which excludes compensation, and the significantly increased number of markers which can be evaluated for each single cell. The Scadden group have recently employed single cell CyTOF to analyse 32 proteins and define 28 subsets of bone marrow stromal cells in homeostatic conditions (Severe *et al*, 2019). Single cell intravital imaging offers new insights into spatial relationships between haematopoietic cells and the microenvironment, showing that leukaemic T-ALL cells are highly motile (Hawkins *et al*, 2016), in contrast with MLL-AF9-driven AML cells where chemoresistance confers an altered migratory phenotype (Duarte *et al*, 2019). By contrast, Long Cai and colleagues have characterised single cells in their spatial environment by computationally integrating data from scRNA-seq and single-molecule fluorescence *in situ* hybridisation (smFISH) (Zhu *et al*, 2018). Integration of scRNA-seq data with ChIP-seq data has been used to infer epigenetic reprogramming and deregulation of TFs in ageing HSCs (Adelman *et al*, 2019), but single cell epigenetic technology has proven more challenging because ChIP-seq is difficult with fewer than 10^3 to 10^4 cells. In this regard, ATAC-seq has emerged as a technology which evaluates chromatin accessibility at the level of the single cell (Buenrostro *et al*, 2015). Finally, CRISPR-seq integrates scRNA-seq with CRISPR-pooled screens to infer causality by gene perturbations at the level of the single cell (Jaitin *et al*, 2016).

Functional characterisation of cells undergoing scRNA-seq remains another important objective to consider during experimental design, and broadly speaking can be achieved by retrospective or prospective methods. The Göttgens group and others have employed flow cytometry to index sort individual cells and facilitate retrospective analysis after lysing cells to acquire RNA (Schulte *et al*, 2015, Macaulay *et al*, 2016, Psaila *et al*, 2016, Knapp *et al*, 2019). Index sorting has proven a powerful partner for scRNA-seq, for instance offering new insights

into the endothelial-to-haematopoietic switch (Baron *et al*, 2018). Secondly, a library of cellular barcodes can be lentivirally integrated into the genomes of single cells to follow them prospectively during differentiation (Naik *et al*, 2013). Lentiviral insertion of barcodes is non-physiological and may alter the biology of single cells, but prospective viral barcoding has introduced fundamental reconfigurations of the haematopoietic tree, such as Naik *et al*'s proposal of DC lineage imprinting during early haematopoiesis (Naik *et al*, 2013).

The evolution of single cell RNA-seq has led to paradigm shifts in our understanding of haematopoiesis. Gene expression analysis of HSC and progenitor populations at the level of individual cells has suggested that lineage-restricted cells emerge directly from a continuum of low-primed undifferentiated HSPCs (Velten *et al*, 2017). This differs conceptually from the hierarchical and stepwise classical model of haematopoiesis, but it remains challenging to reconcile this transcriptome-driven model with the phenotypic observation that populations of blood cells with distinct functional behaviours are readily identifiable with distinct cell surface markers. It can be hypothesised that transcriptional programs characterise intrinsic cell properties, but that discrete phenotypes result from epigenetic modifications or downstream translational processes.

Interestingly, Belluschi *et al* have employed scRNA-seq to identify lymphoid commitment in the $49f^{+}$ human LT-HSC compartment and observed three distinct transcriptional states, speculating that the continuum of transcriptional states observed in HSCs may reflect a common HSC-wide quiescent profile which disguises underlying discrete transcriptional signatures (Belluschi *et al*, 2018). The classical hierarchical model has now given way to a new paradigm of lineage priming during early haematopoiesis, where phenotypically-defined discrete populations of cells do not represent crossroads of cell fate decision-making, but rather they constitute transient phenotypic meeting points for distinctly-primed cells which are already programmed to travel towards divergent final destinations.

1.6 Thesis aims

The Göttgens group combines wet lab functional studies of perturbations within transcriptional networks with rigorous bioinformatics analysis of gene expression data. The group has already constructed computer models of myeloid leukaemia based on TF interactions (Sive *et al*, 2016).

C/EBPα is a significant player in normal myelopoiesis and *CEBPA* is commonly mutated in AML. Furthermore, it interacts significantly with other TFs and is not currently characterised in our group's GRN models. N321D is a C-terminal mutation of *CEBPA* identified in human patients with AML (Papaemmanuil *et al*, 2016) and described as highly lethal in mouse models (Togami *et al*, 2015), and therefore may provide an efficient model to evaluate the effects of *CEBPA* mutation on myeloid transcription. Hoxb8-FL cells are a robust LMPP-like cell line which lies upstream of the CMP to GMP junction, and therefore represent a cellular model for studying pre-leukaemic changes induced by mutant *CEBPA*. *CEBPA*-mutated AML has been extensively studied in cellular and mouse models, however there is comparatively poor understanding of how mutant *CEBPA* is involved in early leukaemogenesis. In summary, our model is designed to characterise pre-leukaemic changes caused by mutant *CEBPA*, on the basis that leukaemia-initiating cells are found downstream of HSCs in *CEBPA* mutations (Kirstetter *et al*, 2008), and that the N321D mutation has been shown to recapitulate AML when retrovirally transfected into BM mononuclear cells (Togami *et al*, 2015).

In this context, my thesis has the following aims:

- Establish a cellular model of *CEBPA*-mutated pre-leukaemia using EV and *CEBPA* WT comparators, by transducing Hoxb8-FL cell lines with retroviral gene expression constructs, and culturing them in cytokine conditions designed to recapitulate early haematopoietic differentiation;

- Analyse the gene expression profile of *CEBPA*-mutated pre-leukaemia at the level of the single cell, aiming to characterise rare transcriptional profiles which cannot be visualised by bulk RNA-seq;

- Examine the genome-wide TF binding and histone modification profile of the *CEBPA* N321D mutation, and correlate this with single cell RNA-seq data to identify candidate genes which may play significant roles in pre-leukaemic changes caused by mutant *CEBPA*, including genes which represent putative candidates for novel therapeutics.

2. Materials and Methods

Unless otherwise stated, all experiments were performed by the author of this thesis.

2.1 Molecular biology

2.1.1 Agarose gel electrophoresis

DNA fragments were separated by electrophoresis in 1 to 1.5% agarose gel. Agarose (Sigma) was prepared in 1 x Tris-Borate EDTA (TBE) electrophoresis buffer and Ethidium Bromide (100 ng/ml) was added as intercalating agent. DNA was mixed with gel loading buffer (bromophenol blue, xylene cyanol, and glycerol) and loaded into wells alongside DNA Hyperladder II (Bioline). Electrophoresis was carried out at 80 to 100 volts (Bio-Rad PowerPac) until DNA had migrated a sufficient distance and fragments were cut out under UV transillumination.

2.1.2 DNA gel extraction

DNA was extracted from agarose gel using the QIAquick Gel Extraction Kit (Qiagen) according to manufacturer's instructions and eluted in 30 to 50 µl elution buffer.

2.1.3 DNA PCR purification

PCR fragments were purified using the QIAquick PCR Purification Kit (Qiagen) according to manufacturer's instructions and eluted in 30 to 50 µl elution buffer.

2.1.4 DNA restriction

Restriction enzymes and buffers from New England BioLabs were used for all digestion reactions, according to manufacturer's instructions. Digests were performed for 1 to 4 hours at 37°C.

2.1.5 Dephosphorylation of DNA fragments

Antarctic phosphatase enzyme and buffer (New England BioLabs) were used to remove the 5' phosphate of DNA restriction fragments to prevent self-ligation, as per manufacturer's instructions.

2.1.6 Blunting of single-stranded overhands in restriction digests

Where required, T4 polymerase (New England BioLabs) was employed to fill in and blunt restriction digest overhangs, as per manufacturer's instructions.

2.1.7 Ligation of DNA fragments

The Promega website http://www.promega.com/a/apps/biomath was used to calculate molar ratio of DNA Insert and plasmid vector, and ligation was performed using the T4 DNA ligase enzyme (New England BioLabs) as per manufacturer's instructions at 16°C overnight.

2.1.8 Preparation of electro-competent DH10β E. Coli

DH10β *Escherichia coli* was streaked out onto Luria Bertani (LB) agar without antibiotics and incubated at 37°C overnight. One colony was then picked into 5 ml of LB broth without antibiotics and incubated at 37°C overnight with shaking. The 5 ml was then inoculated into 500 ml of LB without antibiotics and incubated again at 37°C with shaking for 2 to 3 hour until Optical Density (OD600) reached 0.5. The bacterial culture was then divided into pre-cooled 50 ml falcon tubes, and bacteria was harvested by centrifuging at 3200 x g for 10 minutes at 4°C. The supernatant was discarded and the bacteria were resuspended in ice-cold ultra-pure water and centrifuged twice. Finally, bacteria were resuspended in water, 10% glycerol and 1 mM HEPES pH 8.0, made up into 50 µl aliquots, snap-frozen in liquid nitrogen and stored at -80°C.

2.1.9 DH10β E. Coli transformation

50 µl of Electro-competent DH10β *E. coli* were first thawed on ice, then 2 µl of plasmid DNA was added, and the transformation mixture was placed in a 2 mm cuvette before electroporation at 2.5 kilovolts (25 µF, 250 Ω) on the MicroPulser Electroporator (Bio-Rad). The mixture was then washed with 300 µl of 2 x Tryptone Yeast extract and incubated at 37°C for 45 minutes on a thermo-shaker. Transformed bacteria were plated on LB agar with kanamycin at 37°C overnight.

2.1.10 Plasmid purification by maxi prep

Single bacterial colonies were inoculated into 100 ml of LB with 100 µg/ml of ampicillin, then incubated overnight at 37°C and shaking at 200 to 220 rpm. Maxi prep was then performed

using the Plasmid DNA Purification Kit (Macherey-Nagel), as per manufacturer's instructions, and the final DNA pellet was dissolved in 1 x TE buffer or sterile water.

2.1.11 Plasmid purification by mini prep

Alkaline lysis mini prep was performed to isolate small quantities of plasmid DNA. Single colonies of transformed bacteria were inoculated into 2 ml of LB medium with the appropriate antibiotic, and then incubated at 37°C overnight with vigorous shaking. The culture was then transferred to 1.5 ml Eppendorf tubes, centrifuged at 16,000 x g for 60 seconds, before removing the supernatant. The pellet was then processed with S1 Resuspension buffer, S2 Lysis buffer and S3 Quenching buffer from the NucleoBond Xtra Maxi Kit (Macherey-Nagel), before centrifuging the plasmid DNA in the supernatant at 16,000 x g for 10 minutes and then transferring to a fresh tube. 1 volume of isopropanol was added to precipitate nucleic acids from the supernatant, the solution was vortexed and allowed to stand at room temperature for 2 minutes, before centrifuging at 16,000 x g for ten minutes, washing once with 70% ethanol, and leaving the pellet to air dry. Finally the purified DNA was resuspended in 20 µl of ultra-pure water.

2.1.12 DNA sequencing

The Sanger method was used for DNA sequencing, and performed by Source Bioscience.

2.2 Cloning

Cloning was performed by the author of this PhD and S Kinston.

2.2.1 Original *CEBPA* N321D Plasmid construction

The construct containing the *CEBPA* N321D mutation was assembled from synthetic oligonucleotides and provided by GeneArt (Life Technologies) as an insert into a kanamycin-resistant vector with restriction sites for BlpI and BstAPI (pMK-T CEBPalpha N321D).

2.2.2 Original *CEBPA* N321D Plasmid sequencing

Sanger sequencing was carried out by GeneService (Bioscience) and analysed using SerialCloner software to verify the nucleotide sequence of the expression construct. The following primers were used for sequencing:

Table 2.1 Primers for original *CEBPA* N321D plasmid

Forward primer	5' TCAGAAACAACTTTGGCACA 3'
Reverse primer	5' GTTATCCTCCTCGCCCTTGCTC 3'

2.2.3 *CEBPA* wild-type (WT) plasmid

pMK-T CEBPalpha N321D construct was digested with BlpI and BstAPI restriction enzymes to release the 52 bp region containing the *CEBPA* N321D mutation. The -3368 bp fragment containing the vector and most of *CEBPA* was gel extracted using the QIAquick Gel Extraction Kit (Qiagen), as per manufacturer's instructions.

Complementary primers (Table 2.2) were then used to replace the region containing the mutation with the WT version of the gene. The primers were denatured at 100°C for 10 minutes and then left to cool slowly to room temperature in the presence of annealing buffer (10 mM Tris, 50 mM NaCl, 1 mM EDTA) for them to reanneal. The annealed product was then ligated to the gel extracted band using Gibson assembly cloning strategy with 0.2 U of T5 exonuclease (Epicentre, T5E4111K), 12.5 U of Phusion polymerase (NEB, M0530S), 2000 U of Taq Ligase (NEB, M0208S), and 100 µl of 5X Isothermal (ISO) reaction buffer.

Table 2.2 Primers for Gibson assembly of *CEBPA* WT plasmid

Primer	Sequence
CEBPa Repair oligo FW	AGCAGCGCAACGTGGAGACGCAGCAGAAGGTGCTGGAGCTGACCAGTGAC **AAT**GACCGCCTGCGCAAGCGGGTGGAACAGCTGAGCCGCGAACTGGACAC GCTGCGGG
CEBPa Repair oligo REV	CCCGCAGCGTGTCCAGTTCGCGGCTCAGCTGTTCCACCCGCTTGCGCAGGC GGTC**ATT**GTCACTGGTCAGCTCCAGCACCTTCTGCTGCGTCTCCACGTTGCG CTGCT

Resulting ligations were then transformed and purified by mini prep, and the presence of the wildtype version of the gene was verified by Sanger Sequencing.

2.2.4 Construction of EV, *CEBPA* WT and *CEBPA* N321D retroviral vectors

pMSCV-MLL-ENL-IRES-eGFP retroviral vector was obtained from S Basilico (Göttgens laboratory) and digested with Sal I and Bgl II restriction enzymes. This released MLL-ENL and generated a backbone fragment of 6106 bp that contained pMSCV-IRES-eGFP and was

isolated with the QIAquick Gel Extraction Kit (Qiagen). The ends of this fragment were then blunted and re-ligated to generate the control vector or “Empty Vector” (EV).

The *CEBPA* WT and *CEBPA* N321D insert cDNAs were amplified by PCR using KAPA HiFi Hotstart mix. To this end, we designed a forward oligo that contained the recognition sequence for BamHI followed by the codon of a methionine, in-frame FLAG tag and *CEBPA* (including the methionine codon). The reverse primer contained the sequence of the last coding codon of *CEBPA*. Of note, both primers contained 30 nucleotides of homology with the backbone vector to allow for Gibson assembly cloning strategy, including the region upstream of BamHI site for the forward primer and the IRES region for the reverse primer. The following primers were used:

Table 2.3 PCR primers and conditions to add FLAG tag and Sal I / Bgl II restriction sites to *CEBPA* inserts

Primer	Sequence
SalI_Flg_BmHI_CBPa	TTCTCTAGGCGCCGGAATTCACGCGTCGACGCCACCATGGACTACAAGGACGACGATGACAAGATGGAGTCGGCCGACTTCT
CBPa_SpeI_BGLII_vir	GCGGAATTGGATGCATGGGGTCTTAGATCTTCACGCGCAGTTGCCCATGGC

Stage	**Temperature**	**Time**	**Cycles**
Denaturing	98°C	3 minutes	1
Annealing	98°C	5 seconds	10
	60°C	15 seconds	
	72°C	30 seconds	
Extension	98°C	5 seconds	25
	72°C	60 seconds	
	72°C	5 minutes	
	16°C	Hold	-

The *CEBPA* WT and N321D inserts with FLAG tags and restriction sites were then isolated by gel electrophoresis, joined to the pMSCV-IRES-eGFP control vector by Gibson assembly, transformed into electrocompetent DH-10β cells, and ampicillin-resistant colonies were selected by growth on agar with ampicillin. Plasmid DNA was then isolated by Alkaline Lysis mini prep, and the following primers were used to sequence junctions for the *CEBPA* inserts in the new retroviral vectors and for the IRES sequence in the Empty Vector:

Table 2.4 Primers to sequence ligation junctions in EV, *CEBPA* WT and *CEBPA* N321D retroviral vectors

Primer	Sequence
Cebpa junction Forward primer	CAGACCACCATGCACCT
Cebpa junction Reverse primer	GTCGATGTAGGCGCTGATG
IRES Reverse	AGGAACTGCTTCCTTCACGA

In addition, a number of primers (Table 2.5) were required for Sanger Sequencing along the length of the *CEBPA* WT and *CEBPA* N321D inserts, to ensure that plasmid DNA was composed of the correct and complete nucleotide sequence for the wildtype and mutant inserts.

Table 2.5 Primers for Sanger sequencing along the length of *CEBPA* WT and *CEBPA* N321D inserts

Primer	Sequence	Position
Forward primer	GGCGGCGACTTTGACTAC	344
Forward primer	CAGCAGGAGAAGGCCAAG	322
Forward primer	TACCTGGACGGCAGGCT	475
Reverse primer	CAGGTGCATGGTGGTCTG	683
Reverse primer	GTGCGCGATCTGGAACT	706

The completed vector map is shown in Appendix Figure 1.

2.3 Cell culture

2.3.1 Conditions

Cells were incubated at 37°C in 5% CO_2 and cell culture was performed using standard aseptic technique in a BioMAT2 class II laminar flow cabinet with unidirectional flow of HEPA-filtered air.

2.3.2 B16 cell line

B16 cells were thawed and seeded into 10 cm dishes with 10 ml culture media: RPMI 1640 (Sigma R8758) with 10% foetal bovine serum (FBS), 0.1% 2-mercaptoethanol, 1% Penicillin/Streptomycin and 1% Glutamine. Cells were cultured for 2 to 3 days, then transferred to 175 cm^2 flasks and incubated until confluent. Flt3L conditional media was harvested from the supernatant by pooling supernatant, centrifuging and passing through a 0.2 µM filter, before freezing at -20°C.

2.3.3 293T cell line

293T cells (Graham *et al*, 1977) were cultured in Dulbecco's Modified Eagle's Medium DMEM (Sigma Aldrich D6429), 10% FBS and 1% Penicillin/Streptomycin. These cells were adherent and grown in a monolayer in 10 cm cell culture dishes. They were subcultured and dissociated with Trypsin EDTA (Sigma T3924) in log phase every 2 to 3 days to maintain 30% to 90% confluence.

2.3.4 Hoxb8-FL cell line

The murine cell line Hoxb8-FL (Redecke *et al*, 2013) was kindly donated by Hans Häcker as mouse bone marrow cells infected with ERHBD-Hoxb8 virus. Cells were cultured in RPMI 1640 medium (Sigma R8758) supplemented with 10% Hyclone FBS, 5% Flt3L conditioned media, 1% Penicillin/Streptomycin, 2 mM L-glutamine, 50 mM 2-mercapto-ethanol, and 1 µM β-estradiol (henceforth termed 'Hoxb8 media F+e+').

Cells in full media had lymphoblast-like spherical cell morphology, were non-adherent and could be grown in suspension within sterile tissue culture treated flasks with gas-permeable caps. Cells were subcultured in fresh medium every 2 to 3 days to maintain a concentration of 0.3 to 1.2 x 10^6 cells/ml.

2.3.5 Parental EV, *CEBPA* WT and *CEBPA* N321D cell lines

The Hoxb8-FL cell line was infected with one of the following three retroviral plasmids to generate three cell lines: (i) pMSCV-IRES-eGFP which carries no human transgene, (ii) pMSCV-CEBPAWT-IRES-eGFP, and (iii) pMSCV-CEBPAN321D-IRES-eGFP. For experiments requiring transduced cells only, transduction was followed by cell sorting (Flow Facility, CIMR, Cambridge) to select DAPI negative and GFP^+ cells only.

2.4 Cell biology

Retroviral transduction, differentiation and flow cytometry were performed by the author of this PhD. Cell sorting was performed by the CIMR Flow team and Smart-seq2 protocol was performed by I Kucinski. ChIP-seq was performed by the author of this study and F Calero-Nieto.

2.4.1 Retrovirus production by transfection of 293T cells

2 x 10^6 293T cells were plated in a 10 cm dish at 16 to 20 hours before transfection, aiming for 70% to 80% confluence at time of transfection. 5 µg of plasmid DNA and 5 µg of Psi packaging vector (obtained from the Huntly group) was made up to 10 µl with H_2O and added to 500 µl of plain DMEM, before adding drop-wise to 30 µl of TransIT-LT1 Transfection Reagent (Mirius MIR2300). This was incubated at room temperature for 30 minutes, before distributing drop-wise and evenly over 10 cm dishes containing 293T cells. The transfection mixture was incubated for 24 hours at 37°C in 5% CO_2 and culture medium was then replaced with fresh medium (5 ml RPMI, 10% Hyclone FBS, 1% Penicillin/Streptomycin, 1% glutamine). Retroviral supernatant was recovered after a further 24 hours of culture and passed through a 0.45 µm filter to remove cellular debris, fresh medium was added to the 293T cells and supernatant was recovered again after a further 24 hours.

2.4.2 Retroviral transduction of Hoxb8-FL cells

In each well of a 6-well plate, 1.5 x 10^6 Hoxb8-FL parental cells were placed in 2 ml Hoxb8 media F+e+ and 24 µl polybrene (Sigma) was added to a final concentration of 8 µg/ml, before adding 1 ml of retroviral supernatant. The 6-well plate was then centrifuged at 779 x g for 90 minutes at 32°C, then incubated for 90 minutes at 37°C in 5% CO_2. 1.5 ml of media was removed carefully to avoid disturbing transduced cells lying at the bottom of the wells, and replaced with 2.5 ml of fresh Hoxb8 media F+e+ before incubating overnight at 37°C in 5% CO_2. 2 ml of Hoxb8 media F+e+ was added at 24 hours post-transduction, and cells were assessed for GFP expression at 48 hours post-transduction.

2.4.3 Differentiation assays

EV, *CEBPA* WT or *CEBPA* N321D-transduced cells were cultured for 7 days in Hoxb8 media F+e+, then washed twice in PBS, and resuspended at a concentration of 0.1 x 10^6 cells/ml in differentiation medium.

Differentiation medium was prepared by adding appropriate growth factors to RPMI 1640 medium (Sigma R8758) supplemented with 10% Hyclone FBS, 1% Penicillin/Streptomycin, 2 mM L-glutamine, and 50 mM 2-mercapto-ethanol. The growth factors consisted of either 5% Flt3L conditional medium, 7 ng/ml murine granulocyte-macrophage colony stimulating factor (GM-CSF) (PeproTech), 10 ng/ml murine macrophage colony stimulating factor (M-CSF) (PeproTech), or 10 ng/ml murine IL-3 (PeproTech). Cells were cultured in 6-well plates at 37°C in 5% CO_2. Differentiated cells were adherent and therefore dissociated enzymatically with trypsin-EDTA (Sigma T3924) prior to further processing.

2.4.4 Cytospin and Romanowsky staining

Cytospin was performed by loading Cytofunnels with 50 µl to 100 µl of cells in liquid culture and centrifuging for 2 minutes at 700 rpm on the Shandon Cytospin3, depositing monolayers of cells on Cytoslides. Cells were stained by using a Rapid Romanowsky Stain Pack as per protocol (TCS Biosciences, HS705). Microscopic photographs were taken using Zen Pro software on Axio Imager.Z2 (Zeiss).

2.4.5 Murine haematopoietic progenitor cell staining

Cell staining was performed by using either a Basic Panel (Table 2.6) or an Advanced Panel (Table 2.7). Cells were centrifuged at 300 x g and washed in PBS, then centrifuged again and Fc-blocked using purified anti-CD16/32 antibody (BioLegend). In the case of the Advanced Panel, Fc-block was made up in BD Horizon Brilliant Stain Buffer to prevent staining artefacts caused by the presence of multiple BD Horizon Brilliant fluorescent polymer dyes. Fc-block was carried out for 5 minutes at room temperature, and cells were then conjugated with fluorophore-labelled antibodies in FACS buffer (PBS with 2% Hyclone FBS) for 30 minutes at 4°C.

Cells were washed twice with FACS buffer and then resuspended in 500 µl of FACS buffer with 1 µg/ml of 4',6-diamidino-2-phenylindole (DAPI) or 1:400 of 7-Aminoactinomycin D (7-AAD)

viability dyes for the Basic and Advanced Panels respectively. OneComp compensation beads (Thermo Fisher) were processed alongside cells for use as compensation controls.

Flow cytometry analysis was carried out on the BD LSRFortessa Cell Analyzer (BD Biosciences), and flow cytometry data was processed using FlowJo software.

Table 2.6 Basic Panel of antibodies used for initial flow cytometry and for FACS of mouse cells

Antibody	Clone	Supplier	Catalogue	Dilution
CD 16/32 (Fc block)	93	Invitrogen	14-0161-85	1:100
c-kit (APC/Cy7)	2B8	BioLegend	105826	1:100
Ly-6G (PB)	1A8	BioLegend	127612	1:100
B220 (APC)	RA3-6B2	BioLegend	103212	1:400
CD11b / MAC-1 (PE)	M1/70	BD Pharmingen	553311	1:200

Table 2.7 Advanced Panel of antibodies used for further flow cytometry of mouse cells

Antibody	Clone	Supplier	Catalogue	Dilution
CD 16/32 (Fc block)	93	Invitrogen	14-0161-85	1:100
c-kit (APC/Cy7)	2B8	BioLegend	105826	1:100
Ly-6g (BV510)	1A8	BioLegend	127633	1:100
B220 (AF 700)	RA3-6B2	BioLegend	103232	1:100
CD11b (BV711)	M1/70	BioLegend	101242	1:200
CD11c (PE/Cy7)	N418	BioLegend	117318	1:100
SIRPα (PE)	P84	BD Biosciences	560107	1:100
XCR1 (APC)	ZET	BioLegend	148206	1:100
BST2 (BV421)	927	BD Horizon	566431	1:100
F4/80 (BUV395)	T45-2342	BD Horizon	565614	1:100
MHC class II (PerCP/Cy5.5)	M5/114.15.2	BioLegend	107626	1:400

2.4.6 Murine haematopoietic progenitor cell sorting

Cell sorting was performed on a BD Influx Cell Sorter using BD FACS Software, and cells were sorted into appropriate culture media. For single cell sorting, eight-peak rainbow beads (Spherotech) were employed for laser alignment and BD Accudrop beads were used for calculating drop delay. Single cells were sorted into six standard 96-well PCR plates containing

lysis buffer (SUPERase-In RNase Inhibitor, 10% Triton X-100, Rnase-free H_2O), vortexed for 30 seconds, centrifuged at 700 x g for 60 seconds at 6°C, and then stored at -80°C. Each plate contained thirty-two EV, *CEBPA* WT and *CEBPA* N321D-transduced single cells from a distinct biological replicate at one of two time points.

2.4.7 Single cell RNA sequencing

Single cells were processed by a modified Smart-seq2 protocol. Firstly, the oligo-dT primer was hybridised to the polyA tail of the mRNA molecule by placing 2 µl of the annealing mixture (1 µl free dNTPs (10mM), 0.1 µl tailed oligo-dT oligonucleotides (100 µM), and 0.1 µl ERCCs diluted in 0.8 µl DEPC-treated dH_2O) into each well, centrifuging the plates at 700 x g for 60 seconds, incubating at 72°C for 3 minutes, placing immediately on ice, and then centrifuging again at 700 x g for 60 seconds.

Reverse transcription was then performed by adding 5.7 µl of RT mixture to each well, centrifuging at 700 x g for 60 seconds, and then applying RT thermocycling conditions (Table 2.8). The RT mixture for each well consisted of 0.5 µl Superscript II RT (200 U/µl), Rnase inhibitor (20U/µl), 2.0 µl 5x superscript II first strand buffer, 0.5 µl DTT (100 mM), 2 µl Betaine (5M), 0.06 µl $MgCl_2$ (1M), 0.1 µl TSO (100 µM), and 0.27 µl DEPC-treated dH_2O.

Table 2.8 RT thermocycler conditions

Stage	Temperature	Time	Cycles
RT and template switching	42°C	90 minutes	1
Unfolding of RNA secondary structures	50°C	2 minutes	10
Completion/continuation of RT and template switching	42°C	2 minutes	
Enzyme inactivation	70°C	15 minutes	1
	4°C	Hold	-

Following RT, plates were spun at 700 x g for 60 seconds at 6°C and PCR pre-amplification was carried out by adding 15 µl of PCR mix (12.5 µl of 2x KAPA HiFi Hotstart ReadyMix, 0.25 µl of IS PCR primer (10 µM) and 2.25 µl of nuclease-free dH2O) to each well and applying thermocycler conditions as described in Table 2.9.

Table 2.9 PCR pre-amplification thermocycler conditions

Stage	Temperature	Time	Cycles
Denaturing	98°C	3 minutes	1
Denaturing	98°C	20 seconds	21
Annealing	67°C	15 seconds	
Extension	72°C	6 minutes	
Extension	72°C	5 minutes	1
	4°C	Hold	-

The PCR products were then purified as follows: Ampure XP beads were equilibrated to room temperature for 15 minutes, vortexed, centrifuged at 700 x g for 60 seconds and added to each sample at a volume ratio of 1.0:0.6 for DNA:beads. The mixture was pipetted until homogenous, and transferred to a 96-well plate, then incubated for 8 minutes at room temperature to allow the DNA to bind to the beads. The plates were then placed on a compatible magnetic stand for 5 minutes so that the beads separated from the supernatant, and the liquid carefully removed without disturbing the beads. The beads were then washed twice with 200 µl of freshly made 80% ethanol. The beads were then dried completely, the plates were removed from the magnet stand, and the beads were resuspended in 23 µl of elution buffer before incubating for 2 minutes. The plates were then put back on the magnet for 2 minutes to separate the beads from the solution of eluted PCR products, and 20 µl of the solution was carefully collected without disturbing the beads.

Quality check of the cDNA library was performed by checking size distribution on an Agilent high-sensitivity DNA chip. Quantification was performed using the Scientific Quant-iT™ PicoGreen ® dsDNA Assay Kit.

Library preparation and tagmentation was performed using the Illumina Nextera XT DNA sample preparation kit as per manufacturer's protocol. Samples were diluted to between 0.1 and 0.15 ng/µl cDNA per single cell, 1.25 µl of amplified cDNA was transferred to a new plate, tagmentation mix (2.5 µl Tagmentation DNA buffer, 1.25 µl Amplicon Tagment Mix) was added to each well, plates were centrifuged at 700 x g for 60 seconds, tagmentation reaction was carried out at 55°C for 10 minutes, 1.25 µl of NT buffer was added to each sample to neutralise the tagmentation reaction, and plates were centrifuged again at 700 x g for 60 seconds. Finally, preparation of the sequencing library was completed by adding 3.75 µl of Nextera PCR Master Mix, 1.25 µl of Index Primer 1 (N701-N712) to the corresponding well of

each row of the plate, 1.25 µl of Index Primer 2 (S501-S508) to the corresponding well of each column of the plate, sealing the plate, centrifuging at 700 x g for 60 seconds, and then performing PCR amplification as described in Table 2.10:

Table 2.10 PCR amplification thermocycler conditions

Step	Temperature	Time	Cycles
1	72°C	3 minutes	1
2	95°C	30 seconds	
3	95°C	10 seconds	21
	55°C	30 seconds	
	72°C	60 seconds	
4	72°C	5 minutes	1
5	10°C	Hold	-

Libraries were then pooled from each sample, the total pooled volume was measured, and Ampure XP beads were added firstly in 1.0:0.5 volume ratio, and then in 1.0:0.2 volume ratio for DNA:beads. This double clean-up was designed to remove fragments larger than 600 bp and fragments smaller than 400 bp respectively. The DNA:bead mixture was incubated at room temperature for 5 minutes, placed on a magnetic stand for 2 minutes, and supernatant was removed carefully without disturbing the beads before adding 1ml of freshly prepared 80% ethanol and incubating for 30 seconds on the magnetic stand. Ethanol wash was applied twice, the beads were allowed to dry after each wash, and DNA was then eluted from the beads with elution buffer.

Agilent High-Sensitivity DNA BioAnalyzer chip was employed to check library size distribution between 42 bp and 3561 bp, and the KAPA library quantification kit was used to quantify the pooled library. Libraries were submitted as 10 to 20nM in a volume of at least 15 µl in a 1.5ml Eppendorf. Single cell RNA sequencing was performed by CRUK Cambridge Institute Genomics Core using single-end 50 bp read length, processing 6 plates (576 cells) in 3 lanes on Illumina HiSEq 4000.

2.4.8 ChIP-seq (ChIP, qPCR, library prep and sequencing)

2.4.8.1 Chromatin immunoprecipitation

ChIP nuclei were prepared from EV and *CEBPA* N321D-transduced cells which were collected both before differentiation in Hoxb8 media F+e+ (EV 40 x 10^6 cells, *CEBPA* N321D 80 x 10^6 cells) and after differentiation in Hoxb8 media without estradiol (henceforth 'Hoxb8 media

F+e-') (EV 80 x 10^6 cells, *CEBPA* N321D 160 x 10^6 cells) for 5 days. Cells in F+e+ media were free floating in suspension but differentiated cells in Hoxb8 media F+e- were trypsinised for 1 minute at 37°C before quenching with original media. Cells were centrifuged at 335 x g for 5 minutes at room temperature, cross-linked with formaldehyde 1% (Sigma 47608), stirred with gentle agitation for 10 minutes, and then quenched with 2M glycine. Cells were then stirred for 5 minutes at room temperature, and centrifuged at 300 x g for 5 minutes at 4°C.

Cross-linked cells were then lysed with cell lysis buffer (10 mM Tris pH 8.0, 10 mM NaCl and 0.2% NP-40, and protease inhibitors), nuclei recovered by centrifugation at 600 x g for 5 minutes at 4°C, and nuclei lysates were resuspended in nuclei lysis buffer, protease inhibitors and 10 mM NaBu. Chromatin was then fragmented by sonication on Diagenode Bioruptor UCD-200 for five cycles (30 seconds on, 30 seconds off) before diluting chromatin fragments at a 1:5 ratio of nuclei lysis buffer to IP dilution buffer.

Non-specific chromatin-antibody binding was pre-cleared by incubating in 50 µg of rabbit IgG (Sigma I5006) at 4°C for one hour, then incubating with protein G sepharose beads at 4°C for two hours, permitting separation of IgG-beads complexes from the chromatin in solution.

Precipitation of specific antibody-bound chromatin was then performed: Chromatin from the EV cells was divided into anti-C/EBPα antibody (rabbit polyclonal IgG; Santa Cruz sc-61x), anti-H3K27Ac antibody (rabbit polyclonal IgG; Abcam ab4729), IgG antibody control (rabbit serum; Sigma I5006) and Input sample, whereas *CEBPA* N321D chromatin was divided into anti-C/EBPα antibody (rabbit; Santa Cruz sc-61x), anti-H3K27Ac antibody (rabbit; Abcam ab4729), anti-FLAG M2 antibody (mouse monoclonal IgG; Sigma F3165), IgG antibody control (rabbit) and IgG antibody control (mouse serum; Sigma I5381) and Input sample. The appropriate antibodies were then added (7 µg antibody for TF/IgG and 5 µg antibody for H3K27Ac) to the chromatin, and the samples were incubated on a roller at 4°C overnight. 60 µl of protein A/G sepharose beads were then added per condition, followed by rotation for 2 hours at 4°C. The sepharose / IgG pellets were centrifuged at 5300 x g for 2 minutes, then washed twice with pre-cooled low salt buffer (20 mM Tris pH 8.1, 2 mM EDTA, 150 mM NaCl, 1% Triton X100, 0.01% SDS), once with Lithium Chloride buffer (10 mM Tris pH 8.1, 1 mM EDTA, 0.25 M LiCl, 1% NP40, 1% sodium deoxycholate monohydrate) and twice with 1 x TE buffer.

The DNA-protein-antibody complexes were then eluted from the beads by adding elution buffer (100 mM $NaHCO_3$, 1% SDS) twice in succession and combining the two elutions.

Cross-links were reversed by adding 1 µg RNaseA and 0.3 M NaCl to both immunoprecipitated samples and Input samples, before incubating at 67°C overnight. Endogenous nucleases were then removed by application of 60 µg Proteinase K at 45°C for 2 hours, before purifying DNA with the QIAquick PCR Purification Kit (Qiagen) and eluting in 50 µl elution buffer.

To check that sonication had resulted in correct size distribution of DNA fragments (300 to 500 bp), 0.005 µg of Input DNA was loaded onto the Agilent 2100 BioAnalyzer using the Agilent High Sensitivity DNA Kit as per manufacturer's instructions.

2.4.8.2 Real-time quantitative PCR

Quantitative PCR was performed using Brilliant II SyBr Green QPCR Low ROX Master Mix (Agilent, 600830) on a Stratagene Mx3000P instrument. Input EV F+e+ was used for the standard curve, with dilutions in 1 x TE buffer at 2 ng/µl, 0.4 ng/µl, 0.08 ng/µl, 0.008 ng/µl, and H_2O. Enrichment was assessed at two regulatory elements known to be bound by C/EBPα (*SPI1* -14 and *NFE2*) as well as a negative control region on mouse chromosome 1 (Calero-Nieto *et al*, 2010). ChIP products were diluted to 1:4 in 0.1 x TE buffer. Primers (Table 2.11) and thermocycler conditions (Table 2.12) are given below:

Table 2.11 qPCR primers

	PCR forward primer	PCR reverse primer
SPI1 -14	GCTGTTGGCGTTTTGCAAT	GGCCGGTGCCTGAGAAA
NFE2	TCACACCAGTAGGCAATCCA	GTGGCTAGAGGTGGAACCAG
Negative control region on Chromosome 1	CATAGATGAAGCTGCCACATAGGT	GTGGGCAAGGACAAAGCATTA

Table 2.12 PCR amplification thermocycler conditions

Step	Temperature	Time	Cycles
1	95°C	10 minutes	1
2	95°C	30 seconds	40
	60°C	60 seconds	
	95°C	60 seconds	
3	55°C	30 seconds	1
4	95°C	30 seconds	1
5	4°C	Hold	-

MxPro QPCR software (version 4.10) and Microsoft Excel 2016 were used to analyse data. Fold enrichment levels were calculated relative to IgG.

2.4.8.3 ChIP library preparation and sequencing

Libraries were constructed using TruSeq ChIP Library Preparation Kit Set A and Set B (Illumina, IP-202-1012 and IP-202-1024) to supply 24 indexed Illumina adapters for multiplexed sequencing. Briefly, ChIP enriched DNA was end repaired, purified with the DNA Clean and Concentrator-5 Kit (Zymo Research, D4014) and eluted in 11 µl of elution buffer (Qiagen). A-nucleotide bases were then added to the 3' end of the DNA fragments by incubating with A-tailing mix, before ligating the relevant RNA adapter index and repeating DNA purification. PCR amplification of the elution product was carried out as per Table 2.13:

Table 2.13 PCR amplification thermocycler conditions

Step	Temperature	Time	Cycles
1	98°C	30 seconds	1
2	98°C	10 seconds	18
	60°C	30 seconds	
	72°C	30 seconds	
3	72°C	5 minutes	1
4	4°C	Hold	-

Both the PCR product and 10 µl of 100bp Hyper Ladder II (BioLine) were then run on 2% agarose gel (100 ml 1 x TAE buffer) infiltrated with 10 µl of Sybr Green (Invitrogen), and fragments between 250 and 450 bp were selected, taking care to exclude adapter-dimers at 126 bp. DNA was recovered by using the MiniElute Gel-Extraction Kit (Qiagen) to elute into 11 µl of elution buffer. Fragment size was confirmed on the Agilent 2100 BioAnalyzer using the Agilent High Sensitivity DNA Kit as per manufacturer's instructions. DNA concentration was calculated by using the SYBR FAST Universal qPCR Kit (Kapa Biosystems), diluting in Kapa buffer

at dilutions 1:20, 1:2000, 1:200,000, 1:2000,000 and 1: 200 x 10^6. Libraries were combined into two equimolar mixtures of 45 nM and 38 nM, and sequenced on the Illumina HiSeq 2000 platform at the Genomics Core Facility, Cambridge.

2.5 *In vivo* mouse model

Cell extraction was performed by the author of this PhD and F Calero-Nieto. *In vivo* work was performed by G Giotopoulos (Huntly group).

2.5.1 Mice

Six-week old C57BL/6J mice were housed in a pathogen–free animal facility and were allowed unrestricted access to sterile food and water. Experiments were conducted under a UK Home Office project license, as per UK Home Office regulations. This research has been reviewed by the University of Cambridge Animal Welfare and Ethical Review Body, and is compliant with the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012.

2.5.2 Cell extraction from bone marrow compartment

Bone marrow CD45.2$^+$ 'helper' cells were extracted from tibias, femurs and iliac crests of C57BL/6J mice by making an incision just above the bone marrow compartment and using a 21G needle (BD Microlance 3 0.8 x 40 mm) to flush 5ml of 1 x PBS very slowly through the bone, repeating until the marrow becomes pale. The bone marrow exudate was then passed through a strainer, centrifuged at 300 x g for 7 minutes, incubated with 5 ml of RBC lysis buffer for 10 minutes at room temperature, and centrifuged again at 300 x g for 7 minutes. Cells were then resuspended in 6 ml of 1 x PBS.

2.5.3 Infection with EV and CEBPA N321D-transduced cells

Prior to infection, both cell lines were confirmed negative for mycoplasma by PCR at Surrey Diagnostics. Cells were transplanted via tail vein injection into a total of ten lethally irradiated (10 Gray in split dose) C57BL/6J mice. Five mice were infected with EV-transduced cells and five mice were infected with *CEBPA* N321D-transduced cells. In each case, 5 x 10^5 CD45.1$^+$ GFP$^+$ plasmid-transduced cells and 2.5 x 10^5 CD45.2$^+$ helper cells were transplanted into the CD45.2$^+$ mouse. Peripheral blood was withdrawn by tail vein at weekly intervals for the first four weeks and monthly thereafter. Mice reaching humane end points were killed by Schedule

1 method or by exsanguination under terminal general anaesthetic. In addition to peripheral blood samples *in vivo*, bone marrow and splenic tissue was collected from animals *ex vivo*. Chimerism was evaluated by CD45.1, CD45.2 and GFP markers on flow cytometry.

2.6 Bioinformatics

2.6.1 Analysis of single cell gene expression data

Alignment and mapping of single cell expression data was performed by I Kucinski and E Diamanti. Data was uploaded and analysed on ASAP by the author of this PhD. Further analysis was performed using R language code written by the author with input from W Jawaid.

2.6.1.1 Alignment and mapping

Reads were aligned using GSNAP to the GRCm38 genome and reads spanning exons (ENSEMBL m38.81) were counted using HTSeq. The genome and annotation files were augmented with sequences for External RNA Controls Consortium (ERCC) spike-ins, GFP, human *CEBPA* and pMSCV backbone sequences. It should be noted that alignment to human *CEBPA* did not reliably distinguish the single point mutation between human *CEBPA* WT and human *CEBPA* N321D. Reads were mapped to gene features with HTseqcount.

2.6.1.2 QC

FastQC was then used to apply the following quality control metrics: 200,000 < total reads < 3,000,000; 0.15 < fraction mapped; 1800 < highly expressed genes (defined as counts per million > 50); ERCC fraction < 0.12; and mitochondrial gene fraction < 0.2. Cells which fell outside these parameters were classified as outliers and excluded from further analysis.

2.6.1.3 Data normalisation

For the Automated Single-cell Analysis Pipeline (ASAP), data was normalised using the single-cell Latent Variable Model (Buettner *et al*, 2015). Technical noise was excluded using ERCC spike-ins and the statistical method of Brennecke *et al* (Brennecke *et al*, 2013).

2.6.1.4 Dimensionality reduction

t-distributed Stochastic Neighbour Embedding (tSNE) was performed using ASAP software with Perplexity 25 and Theta 0.5. Results were downloaded from ASAP and plotted in three dimensions using the R package ggplot2.

2.6.1.5 Differential expression

ASAP's DESeq2 function was used to perform differential expression between cell types and conditions on normalized data after filtering.

2.6.1.6 Violin plots

For the generation of Violin Plots, which is not a function provided by ASAP, single cell RNA-seq data was processed on R version 3.5.2 (http://bioconductor/biocLite.R). Briefly, raw counts were parsed in to a dataframe, the dataframe was converted to a matrix, and this was correlated with metadata from cells which passed QC. From the DESeq2 package (version 1.22.2), the function estimateSizeFactorForMatrix was then employed to normalise raw counts. GFP expression data was specifically removed before calculating size factor as it would skew normalisation. Normalised counts were then log transformed and zero values were filtered out. Violin plots were generated by using the geom_violin function from ggplot2 package.

2.6.1.7 Enrichr

The online gene set enrichment analysis tool Enrichr (Kuleshov *et al*, 2016) was used to analyse gene sets identified by differential expression (https://amp.pharm.mssm.edu/Enrichr/).

2.6.2 Analysis of ChIP-seq data

R Hannah performed bioinformatic analysis of ChIP-seq data including genome alignment, peak calling, and read count quantification. The author of this PhD analysed peaks using UCSC genome browser.

2.6.2.1 Genome alignment and peak calling

Raw data in FASTQ format was run through FASTQC to exclude overrepresented sequences and identify adapters. TRIMGALORE was then used to trim adapter sequences from data. The raw reads were then mapped to the mm10 genome using *bowtie2* to generate a SAM file. The SAM file was then converted to BED format by using the sam2bed program from the SAMTOOLS suite, before formatting and extending reads to 200 base pairs. The BED file was then converted to a BEDGRAPH density profile, before converting this in turn to compressed bigWig format which could be visualised on the UCSC Genome Browser.

Peaks were called by running MACS2 (Zhang *et al*, 2008) on the BED file data, using the IgG track as a control, and employing a spread of p-value stringencies. Tracks for multiple stringencies were then visually inspected on UCSC Genome Browser to identify optimal p values for counting genuine peaks only in each ChIP-seq track. An in-house Perl script was then used to modify the peaks to 400 base pairs centred on peak summits.

2.6.2.2 Read count quantification

ChIP-seq was performed for all conditions on two separate occasions, and BEDTOOLS was used to perform intersects and identify a union of peaks at regions bound in both experiments. Read coverage was calculated in regions of union for each condition, and then normalised for peak length and total read counts per 10 million reads. For H3K27Ac, the entire genome was split into 200 bp windows and read coverage was calculated over each of these regions. Counts were normalised to total read counts per 10 million reads, as well as using the S3norm method (Xiang *et al*, 2018) to normalise signal-noise ratio. Scores were then log2 transformed and plotted using gplots.

2.6.2.3 Peak distribution analysis

An in-house script mapped a peak to a gene if within a known promoter region as defined by the mammalian promoter database, or if directly overlapping a gene body based on Refgene coordinates. For intergenic peaks, the script mapped peaks to the nearest gene within 50 kilobases on either side, so that it was possible for an intergenic peak to map up to two genes.

2.6.3 Analysis of CODEX data

The author of this PhD performed manual curation of the CODEX database, adding samples identified by the CODEX webcrawler and also manually searching GEO datasets. R Hannah then processed this data using the CODEX pipeline (Sanchez-Castillo *et al*, 2015), and performed bioinformatic analysis to construct heat maps for both transcription factor and histone profiles.

2.6.3.1 CODEX pipeline

Briefly, public data was converted from SRA format to FASTQ format by using the SRA toolkit. Quality control was performed by FASTQC, and overrepresented sequences and adapters were removed with TRIMGALORE. *Bowtie2* was used to align all experiments to the same

human reference genome GRCh37 (hg19), and produce a SAM file, then transformed to an intermediate BED file before converting to a bigWig read density profile. Peaks were called using MACS2 software (Zhang *et* al, 2008) for a range of p-value stringencies, at which stage tracks were manually inspected to identify the optimal stringency.

2.6.3.2 CODEX heat maps

For transcription factors, a matrix was constructed from a master list (the union) of peaks between all samples, based on binary presence or absence of each peak in each sample. Each region (row) was then scored across all samples (columns) in the matrix to give coverage scores for how many reads overlapped a given peak region. Regions with entirely zero coverage scores across all samples were removed from the matrix, and remaining regions were normalised per sample to the total read number of that sample. The samples in the heat maps were clustered based on pairwise correlation between binding patterns across these regions.

For histones, instead of a master peak list, the entire genome was divided into 200 bp windows and the S3norm package (Xiang *et al*, 2018) was used to normalise coverage scores for both sequencing depth and signal-to-noise ratio.

Further processing was performed using in-house Perl and Shell scripts and the R package. Hierarchical clustering was performed in R using the log2 of the coverage scores, comparing pairwise the binding of each sample (column) across all given regions (rows) in the matrix and then clustering based on similarity of those patterns. The DICE coefficient was used on the discrete data in the binary TF matrix, and Euclidean distance was used for the continuous data (coverage scores for histone binding). This data was then plotted using R libraries and functions: gplots, cba and heatmap.2.

3. Construction and phenotypic characterisation of a stable retroviral expression system for *CEBPA* WT and *CEBPA* N321D

3.1 Construction of retroviral expression system

My research hypothesis is that the expression of mutant *CEBPA* will cause transcriptional and phenotypic alterations in a cell line which replicates an early stage of haematopoiesis. Therefore I set out to construct and validate a retroviral expression system for both *CEBPA* WT and *CEBPA* N321D.

3.1.1 Empty Vector (EV), *CEBPA* WT and *CEBPA* N321D plasmids

The *CEBPA* N321D mutation was constructed from synthetic oligonucleotides and provided by GeneArt (Life Technologies) as an insert into a pMK-RQ (kanR) vector with restriction sites for BLPI I and BstAP I. The plasmid was transformed into DH10β competent cells, colony expanded on agar with kanamycin, and then inoculated into Luria Bertani broth. DNA was purified, then digested by BLPI I and BstAP I restriction enzymes, followed by Gibson Assembly with forward and reverse primers to substitute a single nucleotide base for *CEBPA* WT. PCR was used to amplify and add restriction sites and FLAG tag to both the *CEBPA* WT and N321D inserts (see Figure 3.1), and the products were then ligated into a pMSCV-neo retroviral vector with an IRES-eGFP insert provided by the Huntly lab. In addition, an Empty Vector (EV) for the MSCV plasmid was established by digesting an existing stable retroviral expression vector (provided by S Basilico) with Sal I and Bgl II restriction enzymes to excise the transgene but preserve GFP, purifying the plasmid backbone, blunting and then annealing the two ends.

The retroviral expression system is a bicistronic vector where C/EBPα is translated using a cap-dependent ribosome scanning mechanism with its own 5'-UTR, and translation of GFP is by way of cap-independent IRES-mediated recruitment of ribosomes (Jang *et al*, 1988, Stoneley *et al*, 2004). This mechanism facilitates co-expression of C/EBPα and GFP, although there is evidence to suggest that IRES-mediated co-expression can vary with cellular context (Mizuguchi *et al*, 2000).

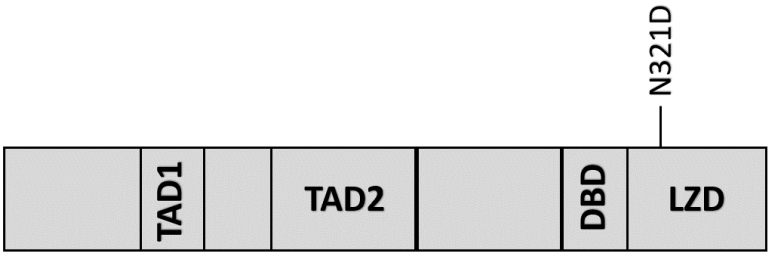
A

ATGGAGTCGGCCGACTTCTACGAGGCGGAGCCGCGGCCCCCGATGAGCAGCCACCTGCAGAGCCCCCCGCACGCGCCCAGCAGCG
CCGCCTTCGGCTTTCCCCGGGGCGCGGGCCCCGCGCAGCCTCCCGCCCCACCTGCCGCCCCGGAGCCGCTGGGCGGCATCTGCGAG
CACGAGACGTCCATCGACATCAGCGCCTACATCGACCCGGCCGCCTTCAACGACGAGTTCCTGGCCGACCTGTTCCAGCACAGCCG
GCAGCAGGAGAAGGCCAAGGCGGCCGTGGGCCCCACGGGCGGCGGCGGCGGCGGCGACTTTGACTACCCGGGCGCGCCCGCGG
GCCCCGGCGGCGCCGTCATGCCCGGGGGAGCGCACGGGCCCCCGCCCGGCTACGGCTGCGCGGCCGCCGGCTACCTGGACGGCA
GGCTGGAGCCCCTGTACGAGCGCGTCGGGGCGCCGGCGCTGCGGCCGCTGGTGATCAAGCAGGAGCCCCGCGAGGAGGATGAA
GCCAAGCAGCTGGCGCTGGCCGGCCTCTTCCCTTACCAGCCGCCGCCGCCGCCGCCCTCGCACCCGCACCCGCACCCGCCGCCC
GCGCACCTGGCCGCCCCGCACCTGCAGTTCCAGATCGCGCACTGCGGCCAGACCACCATGCACCTGCAGCCCGGTCACCCCACGCC
GCCGCCCACGCCCGTGCCCAGCCCGCACCCCGCGCCCGCGCTCGGTGCCGCCGGCCTGCCGGGCCCTGGCAGCGCGCTCAAGGGG
CTGGGCGCCGCGCACCCCGACCTCCGCGCGAGTGGCGGCAGCGGCGCGGGCAAGGCCAAGAAGTCGGTGGACAAGAACAGCAAC
GAGTACCGGGTGCGGCGCGAGCGCAACAACATCGCGGTGCGCAAGAGCCGCGACAAGGCCAAGCAGCGCAACGTGGAGACGCA
GCAGAAGGTGCTGGAGCTGACCAGTGACGATGACCGCCTGCGCAAGCGGGTGGAACAGCTGAGCCGCGAACTGGACACGCTGCG
GGGCATCTTCCGCCAGCTGCCAGAGAGCTCCTTGGTCAAGGCCATGGGCAACTGCGCG

B

MESADFYEAEPRPPMSSHLQSPPHAPSSAAFGFPRGAGPAQPPAPPAAPEPLGGICEHETSIDISAYIDPAAFNDEFLADLFQHSRQQEKA
KAAVGPTGGGGGGDFDYPGAPAGPGGAVMPGGAHGPPPGYGCAAAGYLDGRLEPLYERVGAPALRPLVIKQEPREEDEAKQLALAGL
FPYQPPPPPPPSHPHPHPPPAHLAAPHLQFQIAHCGQTTMHLQPGHPTPPPTPVPSPHPAPALGAAGLPGPGSALKGLGAAHPDLRAS
GGSGAGKAKKSVDKNSNEYRVRRERNNIAVRKSRDKAKQRNVETQQKVLELTSDDDRLRKRVEQLSRELDTLRGIFRQLPESSLVKAMG
NCA

C



D

TTCTCTAGGCGCCGGAATTCACGCGTCGACGCCACCATGGACTACAAGGACGACGATGACAAGATGGAGTCGGCCGACTTCT

E

GCCATGGGCAACTGCGCGTGAAGATCTAAGACCCCATGCATCCAATTCCGc

Figure 3.1 Mutated *CEBPA* in the *CEBPA* N321D cell line. (A) cDNA, and (B) amino acid sequences. In-frame single nucleotide substitution shown after 321st codon, resulting in mutation from asparagine AAT to aspartate **GAT**. (C) Mutation shown in model of C/EBPα protein which characterises TransActivation Domains (TAD) 1 and 2, DNA Binding Domain (DBD) and Leucine Zipper Domain (LZD). (D) shows cDNA sequence of forward oligonucleotide for Gibson PCR assembly of restriction sites and FLAG tags onto *CEBPA* WT and N321D inserts, with yellow vector (Sal I is underlined), FLAG tag no colour and underlined, green *CEBPA*. (E) shows cDNA sequence of reverse oligonucleotide for Gibson PCR with yellow vector (Bgl II is underlined) and green *CEBPA* template sequence.

3.1.2 Verification of integration of *CEBPA* sequences into retroviral model

Successful integration of the *CEBPA* WT and N321D sequences together with FLAG tags was confirmed by Sanger sequencing. Primers were designed specifically to verify *CEBPA* WT and N321D mutation sequences.

3.1.3 Retroviral transduction of Hoxb8-FL cell line

293T cells were transfected with Psi packaging vector and the vector containing the cDNA from either the EV, *CEBPA* WT or *CEBPA* N321D plasmids. The viral supernatant was recovered at 24 hours and 48 hours post-transfection. Supernatant was then transduced into Hoxb8-FL

parental cells to produce cell lines which expressed EV, *CEBPA* WT and *CEBPA* N321D. Together with non-transduced parental control cells, these three cell lines were then examined for variation in cell growth, immunophenotypic profile and morphological appearance under different cytokine conditions.

3.2 Characterisation of transduced cell lines in self-renewal conditions

Hoxb8-FL cells contain a fusion construct of the hormone-binding domain of the estrogen receptor and the N-terminus of Hoxb8 (Wang *et al*, 2006), and in the presence of Flt3L and estradiol are maintained in an early non-differentiated progenitor state with stable growth, conditional immortalisation and multi-lineage potential (Redecke *et al*, 2013). Transduced cell lines were examined in the presence of both Flt3L and estradiol to evaluate if the *CEBPA*-transduction affected Hoxb8-mediated programs of differentiation arrest (Knoepfler *et al*, 2001).

3.2.1 *CEBPA* WT-transduced cells overcome Hoxb8 effect in the presence of estradiol

Expression of myeloid, lymphoid and dendritic cell markers was tested to look for evidence of differentiation within any of the transduced cell lines, using parental Hoxb8-FL cells as a control. The following flow cytometry panel included markers employed by Redecke *et al* in their characterisation of Hoxb-FL cells and additional markers designed to further elucidate dendritic cell (DC) lineage:

Table 3.1 Flow cytometry panel designed to evaluate differentiation of Hoxb8-FL cells and transduced cell lines

CD marker	Cell type
c-kit/CD117	Haematopoietic stem cells and progenitors
Ly6g/Gr-1	Granulocytes
B220/CD45R	B lymphocytes, plasmacytoid dendritic cells
CD11b	Granulocytes, monocytes/macrophages, dendritic cells
CD11c	Dendritic cells
XCR1	Conventional dendritic cell type 1
SIRPα	Conventional dendritic cell type 2
BST2/PDCA1	Plasmacytoid dendritic cells
F4/80	Monocytes/macrophages
MHC class II	Antigen presenting cells

Importantly, it was noted that the parental cells and the EV transduced cells had very similar profiles on flow cytometry in the presence of estradiol, suggesting that the plasmid itself does not have an effect and that effects seen on flow cytometry are due to the *CEBPA* insert.

Interestingly, the *CEBPA* WT-transduced cells demonstrated increased expression of CD11b and SIRPα in the presence of estradiol at Day 2 post-transduction (see Figure 3.2), suggesting that these cells were differentiating down a conventional dendritic cell (cDC) pathway. Expression of *CEBPA* N321D prevented the acquisition of dendritic cell markers. By Day 5, the decreased proportion of *CEBPA* WT-transduced cells made it difficult to continue characterising them by flow cytometry, but no marked difference was seen in any of the other transduced cell lines up to and including Day 18 in the presence of estradiol.

The *CEBPA* WT-transduced cells form two subpopulations: one which is c-kit ^{high} and does not express CD11b or SIRPα, and the other which is c-kit ^{low} and does express CD11b and SIRPα. When these two subpopulations are evaluated separately for GFP expression, it is clear that the first subpopulation expresses less GFP than the second subpopulation (see Figure 3.3). One explanation for this observation is that the *CEBPA* WT-transduced cells which show a dendritic cell phenotype in the presence of estradiol may comprise only those cells which incorporate a higher number of plasmid copy numbers.

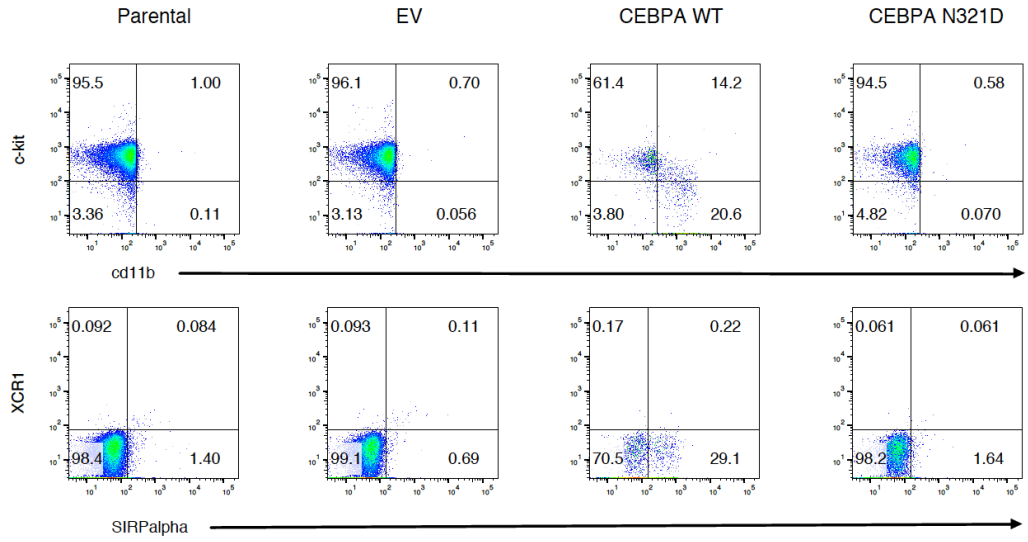


Figure 3.2 Comparison of c-kit, CD11b, XCR1 and SIRPα expression in cell lines. *CEBPA* WT-transduced cells show increased expression of CD11b and SIRPα in F+e+ media following transduction. (3 experimental replicates, 1 replicate showed CD11b only; example shown is representative).

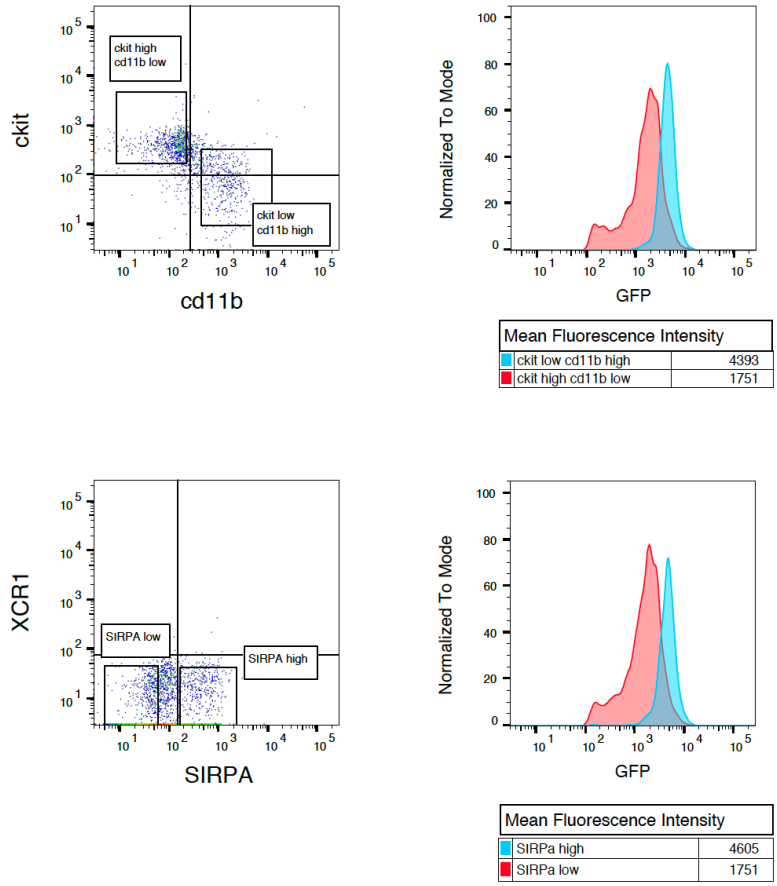


Figure 3.3 Two subpopulations are observed in *CEBPA* WT-transduced cells in the presence of estradiol. Increased GFP expression is seen in cells which are c-kit ^{low} and which express CD11b and SIRPα.

One interesting observation is that the few cells persisting from the *CEBPA* WT-transduced cell line, which expressed CD11b and SIRPα in the presence of Flt3L and estradiol from as soon as Day 2 post-transduction (see Figure 3.2 above), no longer show any significant difference from other cell lines at Day 10 (Figure 3.4). The discrepancy between Day 10 and Day 2 expression profiles of *CEBPA* WT-transduced cells in the presence of estradiol may be explained by the observation at Day 2 (see Figure 3.3) that it is only a subpopulation of *CEBPA* WT-transduced cells expressing more GFP, presumably having incorporated a higher number of *CEBPA* WT plasmid copies, which differentiate into a dendritic cell phenotype. It is possible that these cells have fully differentiated and progressed to cell death by Day 10, leaving behind only Hoxb8-FL cells which incorporated single *CEBPA* WT plasmid copies and which have a similar phenotype to other cell lines in the presence of estradiol.

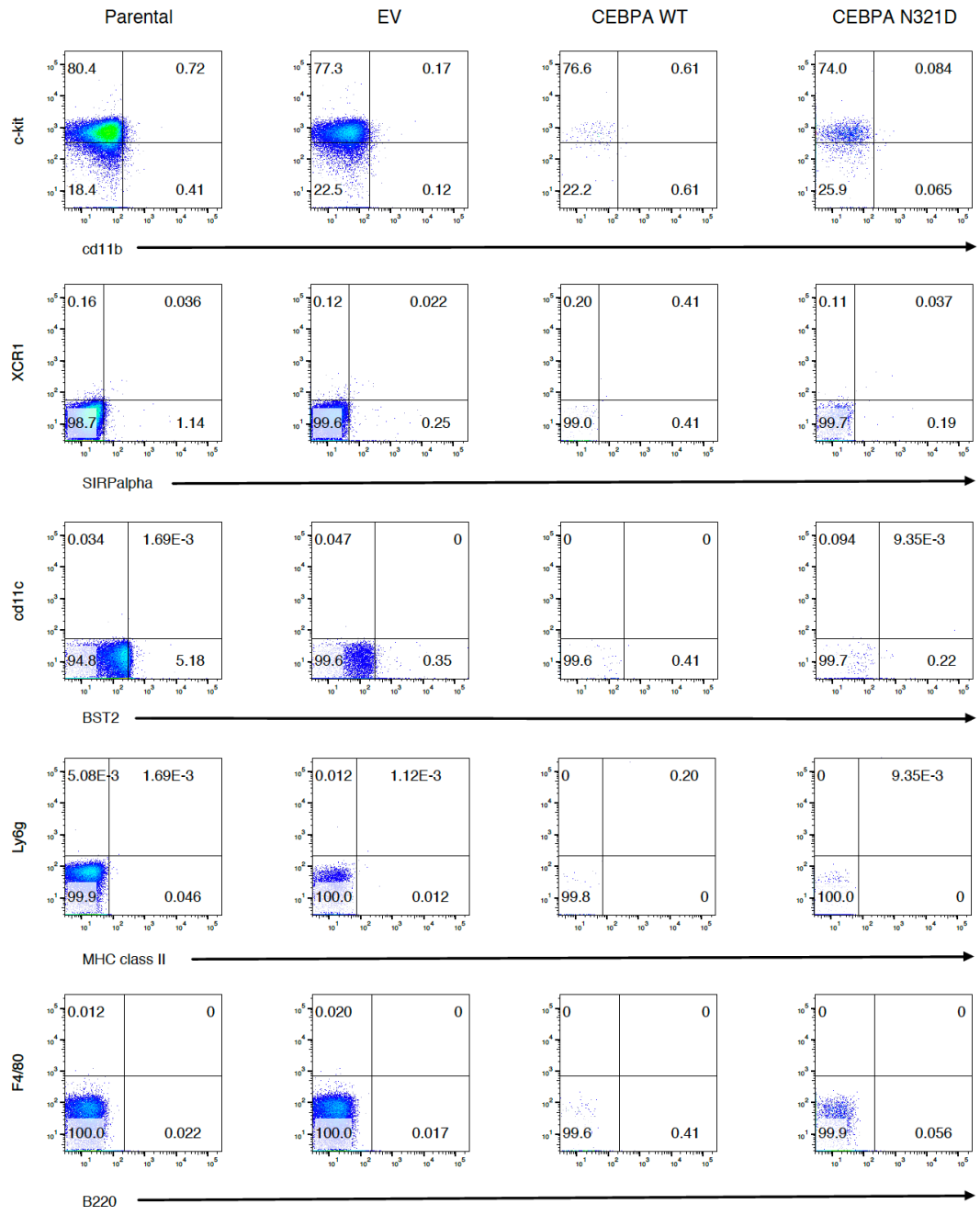
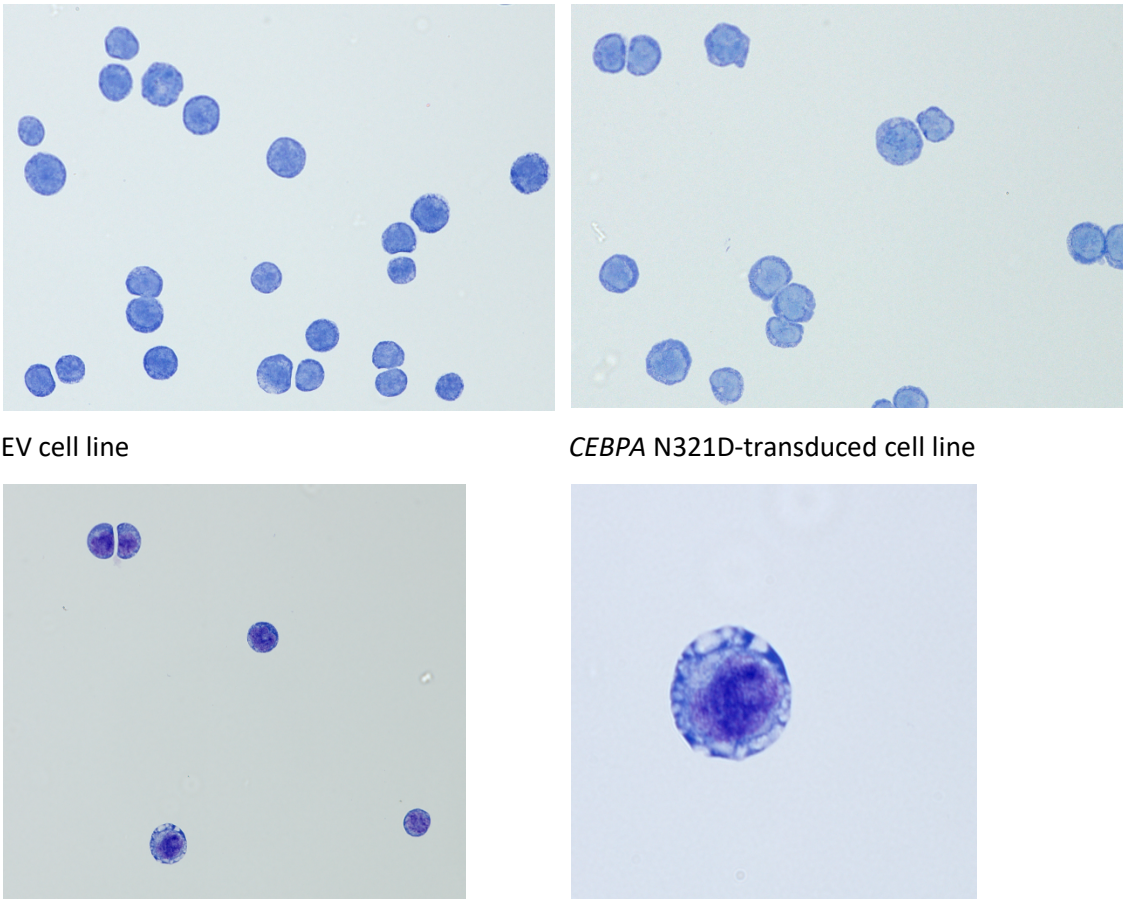


Figure 3.4 Differentiation profile of transduced cell lines at Day 10 in the presence of Flt3L and estradiol. All cell lines are c-kit ^{high} and fail to express lineage markers at Day 10 in the presence of estradiol. (1 replicate for all cell lines, 3 additional replicates showed parental, EV and *CEBPA* N321D only; example shown is representative).

3.2.2 Morphological examination of *CEBPA* WT-transduced cells confirms differentiation in the presence of estradiol

Following the finding that a population of *CEBPA* WT-transduced cells demonstrated a distinct antigenic phenotype on flow cytometry in Flt3L and estradiol, it was important to examine

whether this was reflected by a difference in morphological phenotype. Due to the rapid decrease in *CEBPA* WT-transduced cells as a proportion of total cells after transduction, cell sorting was carried out to isolate GFP$^+$ EV, *CEBPA* WT and *CEBPA* N321D-transduced cells at 4 days post-transduction. Cytospin was performed (span at 700 rpm for 2 minutes, cells were allowed to dry) and then stained using Rapid Romanowsky Stain Pack (TSC Biosciences). Photographs were taken using Axio Imager for bright field microscopy (see Figure 3.5). The EV and *CEBPA* N321D-transduced cell lines showed poorly differentiated cells with immature dispersed chromatin and high nuclear:cytoplasmic ratio reflecting their early haematopoietic progenitor status. By contrast, *CEBPA* WT-transduced cells showed denser clumped chromatin and plentiful vacuoli in the cytoplasm (Figure 3.5).



EV cell line

CEBPA N321D-transduced cell line

CEBPA WT-transduced cell line

Figure 3.5 Morphological appearances of GFP$^+$ cells in presence of Flt3L and estradiol EV and *CEBPA* N321D-transduced cells are poorly differentiated in comparison with mature clumped chromatin and vacuolated cytoplasm in *CEBPA* WT-transduced cells. (1 replicate).

3.2.3 Cell growth suggests that *CEBPA* WT and N321D transduced cells have a competitive disadvantage to EV-transduced Hoxb8-FL cells in the presence of estradiol

Hoxb8-FL cells transduced with EV, *CEBPA* WT and *CEBPA* N321D all continued to grow indefinitely in the presence of Flt3L and estradiol, however the proportional growth rates of the transfected GFP$^+$ compartment showed significant variation between cell lines. Initial transduction rate for EV cells was around 70% at Day 2, and this increased to nearly 90% by Day 20. By contrast, the *CEBPA* WT cells were approximately 10% transduced at Day 2, and this decreased to 0.5% by Day 5 which was then maintained long-term despite expanding the cells every 2 to 3 days. The *CEBPA* N321D cells showed 25% transduction at Day 2, which gradually decreased to 10% by Day 20.

This demonstrated a functional variation between the three cell lines. The EV-transduced cells maintained the Hoxb8-FL growth pattern, whereas the *CEBPA* WT and N321D transduced cells had a competitive disadvantage (see Figure 3.6). This suggests that both *CEBPA* WT and N321D interfere with the differentiation arrest program in Hoxb8-FL parental cells when cultured in self-renewal conditions.

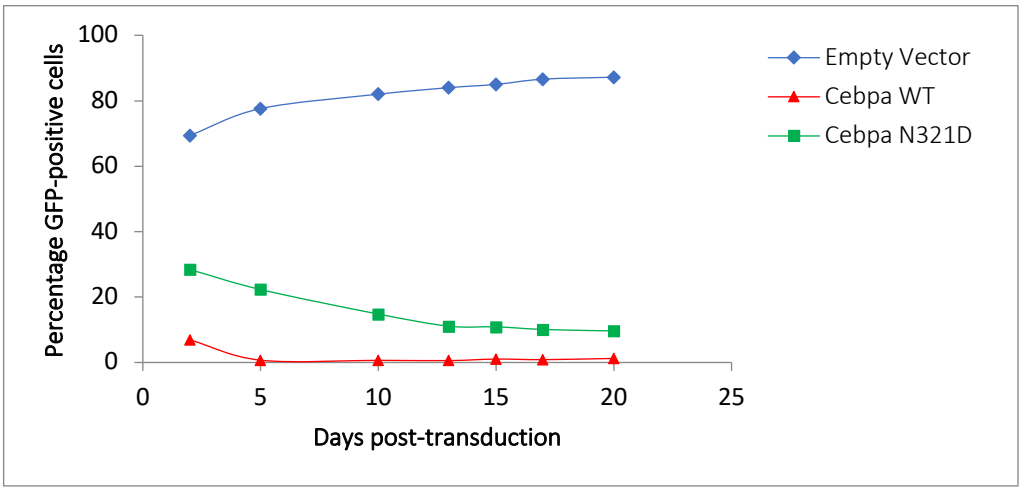


Figure 3.6 Competition assay of Hoxb8-FL parental cells after transduction with retroviral vectors. Decreased proportion of GFP$^+$ *CEBPA* WT and N321D-transduced cells, contrasting with stable growth of EV-transduced cells.

Graphical representation of logarithmic growth in the three transduced cells lines also demonstrated the variation in growth rate within the GFP$^+$ compartment (see Figure 3.7).

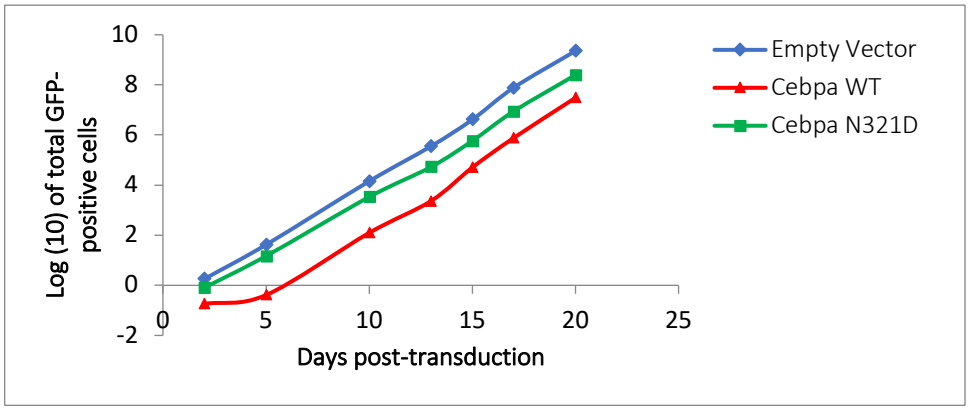


Figure 3.7 Logarithmic growth chart for transduced cell lines. Constant growth rates seen for Hoxb8-parental cells transduced by EV, *CEBPA* WT and *CEBPA* N321D.

3.3 Characterisation of transduced cell lines upon differentiation

Hoxb8-FL cells have multi-lineage potential and can be differentiated into myeloid and lymphoid cell types upon withdrawal of estradiol. Transcriptionally, they are most similar to LMPP cells (Redecke *et al*, 2013). Specifically, they are able to differentiate into cDCs and plasmacytoid dendritic cells (pDCs) in the presence of Flt3L, into DCs and granulocytes in the presence of GM-CSF, and into macrophages in the presence of M-CSF. In addition, it is known that granulopoiesis can be rescued in C/EBPα $^{-/-}$ mice in the presence of exogenous IL-3 (Hirai *et al*, 2006). The EV, *CEBPA* WT and *CEBPA* N321D-transduced cell lines were therefore differentiated in these four cytokine conditions for 5 days.

3.3.1 Flow cytometry characterisation of transduced cell lines upon differentiation in GM-CSF, IL-3, M-CSF, and Flt3L

EV and *CEBPA* N321D-transduced cells maintained in Flt3L and estradiol remained c-kit ^{high} and expressed no lineage markers up to Day 10 (Figure 3.4). This was in keeping with the behaviour of Hoxb8 parental cells which are conditionally immortalised in the presence of estradiol.

Differentiation was then performed by removing estradiol and culturing cells in the presence of GM-CSF, IL-3, M-CSF and Flt3L cytokines, and comparing the flow cytometry profile of GFP^{+} transduced cells with GFP^{-} non-transduced cells (Figure 3.8). This initial experiment was performed with a simplified flow panel using markers of granulocytic (Gr1), dendritic (CD11b), or lymphocytic (B220) differentiation (Figure 3.8). There was no significant phenotypic difference between the three transduced models in GM-CSF, IL-3 or M-CSF, to suggest that *CEBPA* N321D has any substantial effect on granulopoiesis, emergency granulopoiesis or

macrophage development of Hoxb8-FL cells during differentiation. Rather, this initial exploratory data suggested that an important early haematopoietic effect of the *CEBPA* N321D mutation was related to dendritic differentiation pathways, and it was decided to explore this phenotype further by using the flow cytometry panel described in Table 3.1.

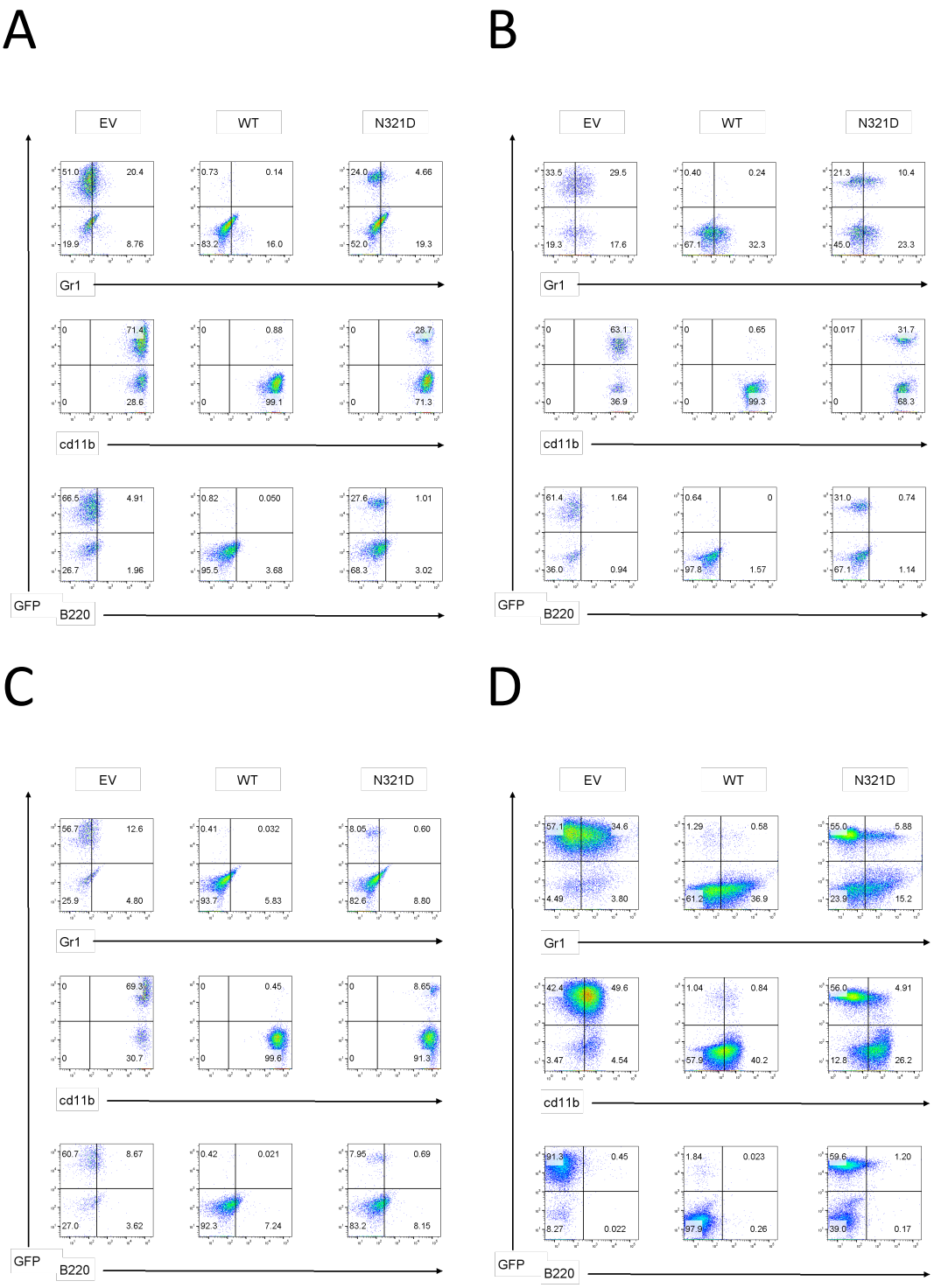


Figure 3.8 Differentiation profile of transduced cell lines at Day 5 in the presence of (A) GM-CSF, (B) IL-3, (C) M-CSF, and (D) Flt3L. *CEBPA* N321D-transduced cells show marked reduction in CD11b expression when cultured in Flt3L. (2 replicates for all cell lines, 1 additional replicate for EV and *CEBPA* N321D only; example is broadly representative).

By Day 5 of differentiation in Flt3L, the EV-transduced cells showed marked expression of CD11b, SIRPα, CD11c and BST2, and showed no expression of XCR1 (Figure 3.9).

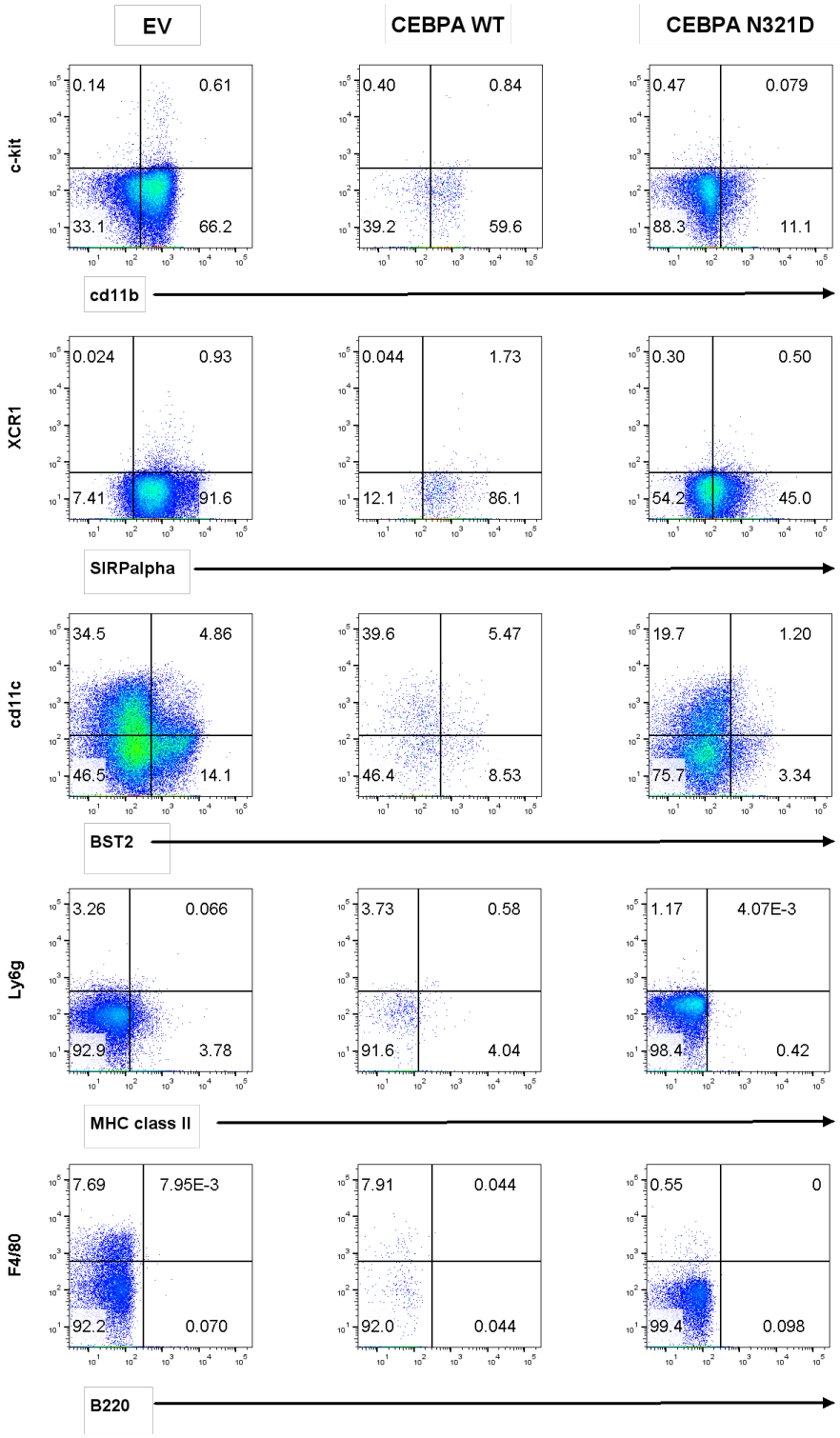


Figure 3.9 Differentiation profile of transduced cell lines at Day 5 in the presence of Flt3L and withdrawal of estradiol. *CEBPA* N321D-transduced cells show decreased expression of CD11b, SIRPα, CD11c and BST2 at Day 5 of differentiation. (1 replicate for all cell lines and full panel as shown, 3 additional replicates for EV and CEBPA N321D cell lines and full marker panel, 2 additional replicates for all cell lines and all cell markers except XCR1).

This confirmed the data demonstrated by Redecke *et al* in their original description of Hoxb8-FL cells, suggesting that this cell line differentiates into both cDCs and pDCs in Flt3L, although our additional characterisation of these cells as SIRPα^{+} XCR1^{-} suggests that they are type 2 cDCs. Interestingly, the *CEBPA* WT-transduced cells showed a similar profile to EV cells after 5 days of differentiation in Flt3L. By contrast, the *CEBPA* N321D-transduced cell line demonstrated reduced expression of both conventional (CD11b, CD11c, SIRPα) and mature plasmacytoid (CD11c, BST2) DC markers, in comparison with EV and *CEBPA* WT-transduced cells (Figure 3.9).

By Day 10, *CEBPA* N321D-transduced cells were also characterised by decreased expression of MHC class II, CD11c, BST2 and F4/80 (Figure 3.10):

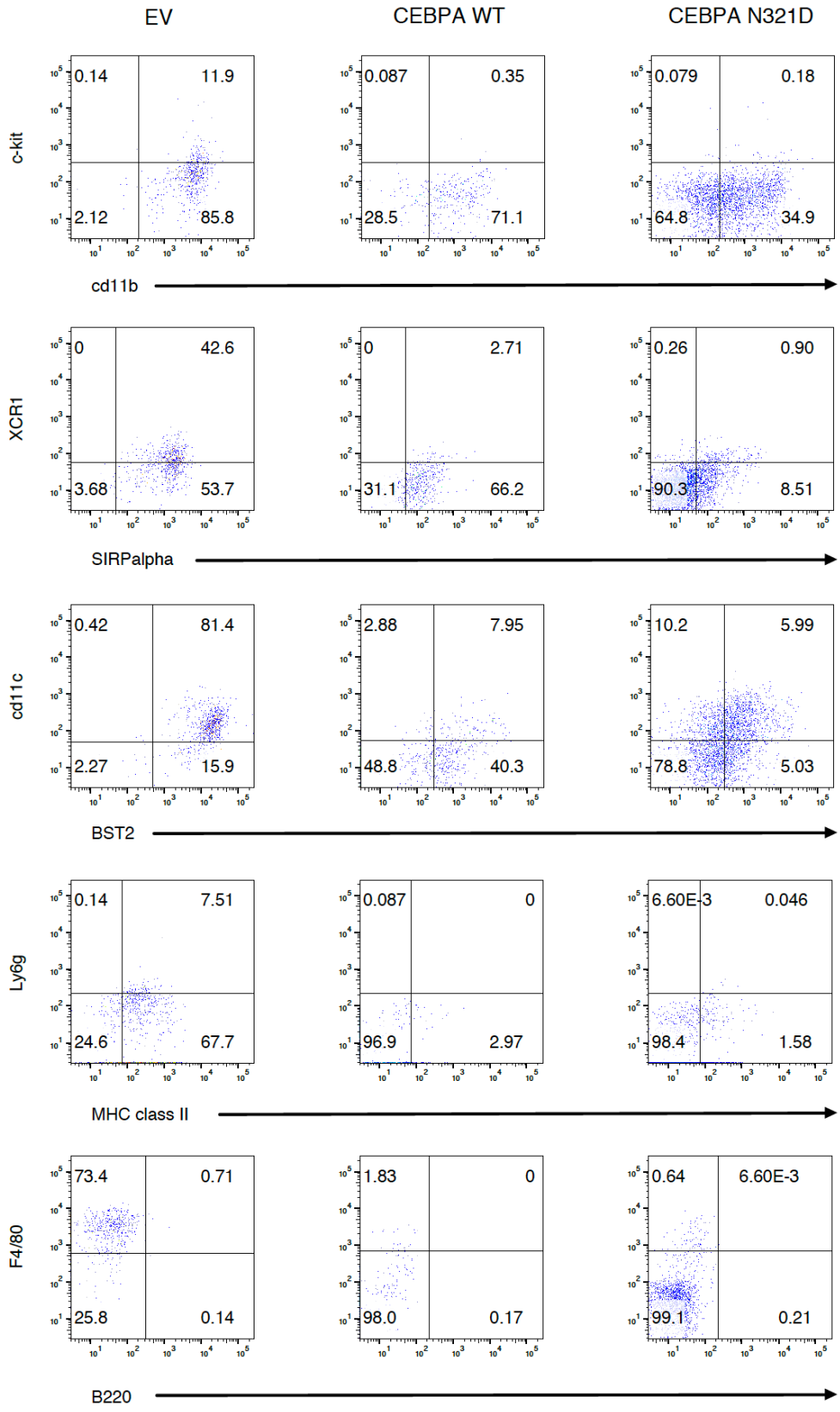


Figure 3.10 Differentiation profile of transduced cell lines at Day 10 in the presence of Flt3L and withdrawal of estradiol. (1 replicate for all cell lines, 3 additional replicates for EV and *CEBPA* N321D cell lines only).

3.3.2 Morphological examination shows that *CEBPA* N321D-transduced cells display differentiation defect compared to EV-transduced cells in the presence of Flt3L

Flow cytometry was repeated on cells differentiated in Flt3L during two separate collections of ChIP nuclei and demonstrated a similar profile as described above. *CEBPA* WT-transduced cells rapidly decreased to less than 0.5% of total cells after differentiation, and therefore morphological examination of GFP^+ sorted cells was performed only on EV and *CEBPA* N321D-transduced cells at Day 5 of differentiation:

A B

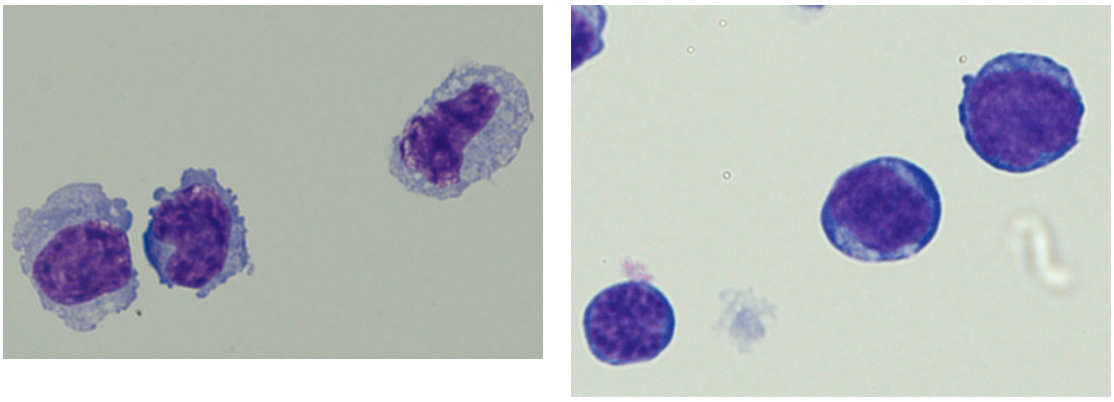


Figure 3.11 Morphological appearances of cell lines after differentiation in Flt3L for 5 days. (A) EV-transduced cell lines are morphologically more differentiated than (B) *CEBPA* N321D-transduced cells after withdrawal of estradiol. (2 replicates, examples shown are broadly representative).

Differentiated EV and *CEBPA* N321D-transduced cells demonstrated a clear morphological difference with differentiation block evident on morphological examination of the *CEBPA* N321D-transduced cells. The EV transduced cells had relatively small nuclei and vacuolated cytoplasm, and some cells suggested pre-DC appearances consisting of lateralised reniform nuclei, typical Golgi regions and a few prominent dendritic projections. By contrast, the *CEBPA* N321D-transduced cells had larger nuclear:cytoplasmic ratio and no evidence of dendritic differentiation (see Figure 3.11).

3.3.3 Cell growth shows that *CEBPA* N321D-transduced cells are immortalised independently of Hoxb8 expression

Unsurprisingly, the distinct differentiation profile of EV and *CEBPA* N321D-transduced cells resulted in very different growth profiles. By Day 12 post-differentiation, not many live viable cells can be observed in the EV cell line. However, *CEBPA* N321D-transduced cells continued

to grow in the absence of estradiol in the media. We kept these cells in culture up to Day 56 without any signs of reduction in growth capabilities (see Figures 3.12 and 3.13).

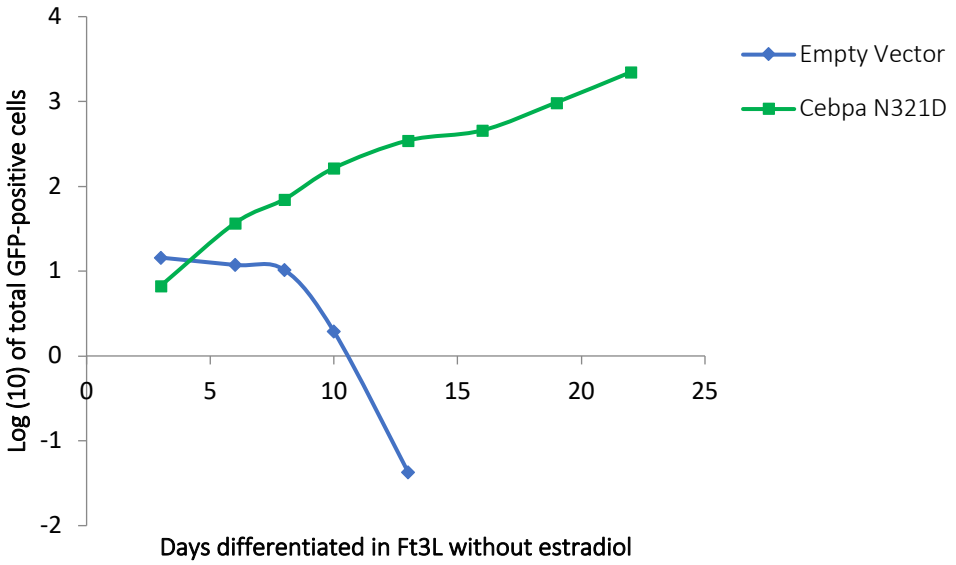


Figure 3.12 Logarithmic growth chart for transduced cell lines. *CEBPA* N321D-transduced cells demonstrate growth advantage on differentiation in Flt3L without estradiol.

3.4 Differentiation of *CEBPA* N321D cell line for extended time period in Flt3L suggests Hoxb8-independent immortalised pDC progenitor phenotype

Having established in multiple transduction experiments that the *CEBPA* N321D-transduced cells under-expressed markers of dendritic cell maturation at Day 5 of differentiation in Flt3L, the next question was to follow mutant transfected cells in Flt3L without estradiol and investigate their growth and phenotype over an extended period of time.

CEBPA N321D-transduced cells in Hoxb8 media F+e- were vigorously pipetted, centrifuged, and cells were then resuspended in fresh F+e- media every 2 to 3 days. Cells continued to grow until the experiment was discontinued at Day 56:

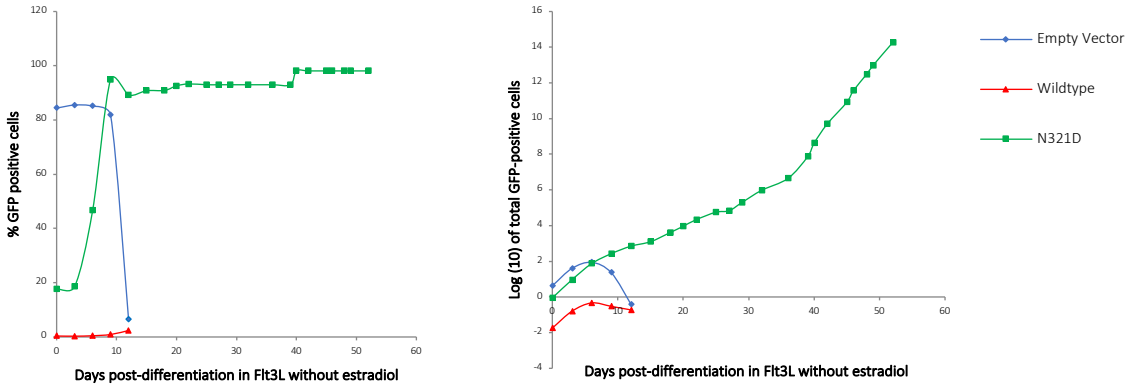


Figure 3.13 Indefinite expansion of *CEBPA* N321D cell line in Flt3L without estradiol. (A) Percentage of GFP^+ cells over time. (B) Log (10) of GFP^+ cells over time.

CEBPA N321D-transduced cells in Flt3L therefore appeared to demonstrate a Hoxb8-independent immortalised phenotype. Morphological appearance suggested differentiation arrest with cytoplasmic vacuoles, Golgi regions, indented and occasionally reniform nuclei with partially condensed chromatin, increased nuclear:cytoplasmic ratio, and small dendritic projections:

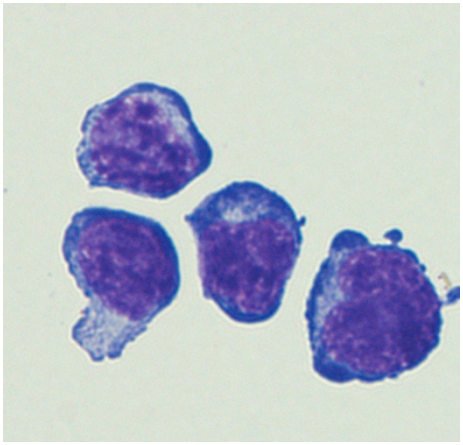


Figure 3.14 *CEBPA* N321D-transduced cells after 56 days of differentiation in Hoxb8 media F+e-. (1 replicate only).

Furthermore, flow cytometry of the immortalised cell line at Day 56 demonstrated a dramatically different phenotype from the phenotype described at Days 5 and 10 above. These cells were c-kit negative, expressed no cDC markers, but did express markers suggestive of a $BST2^-$ pDC progenitor phenotype:

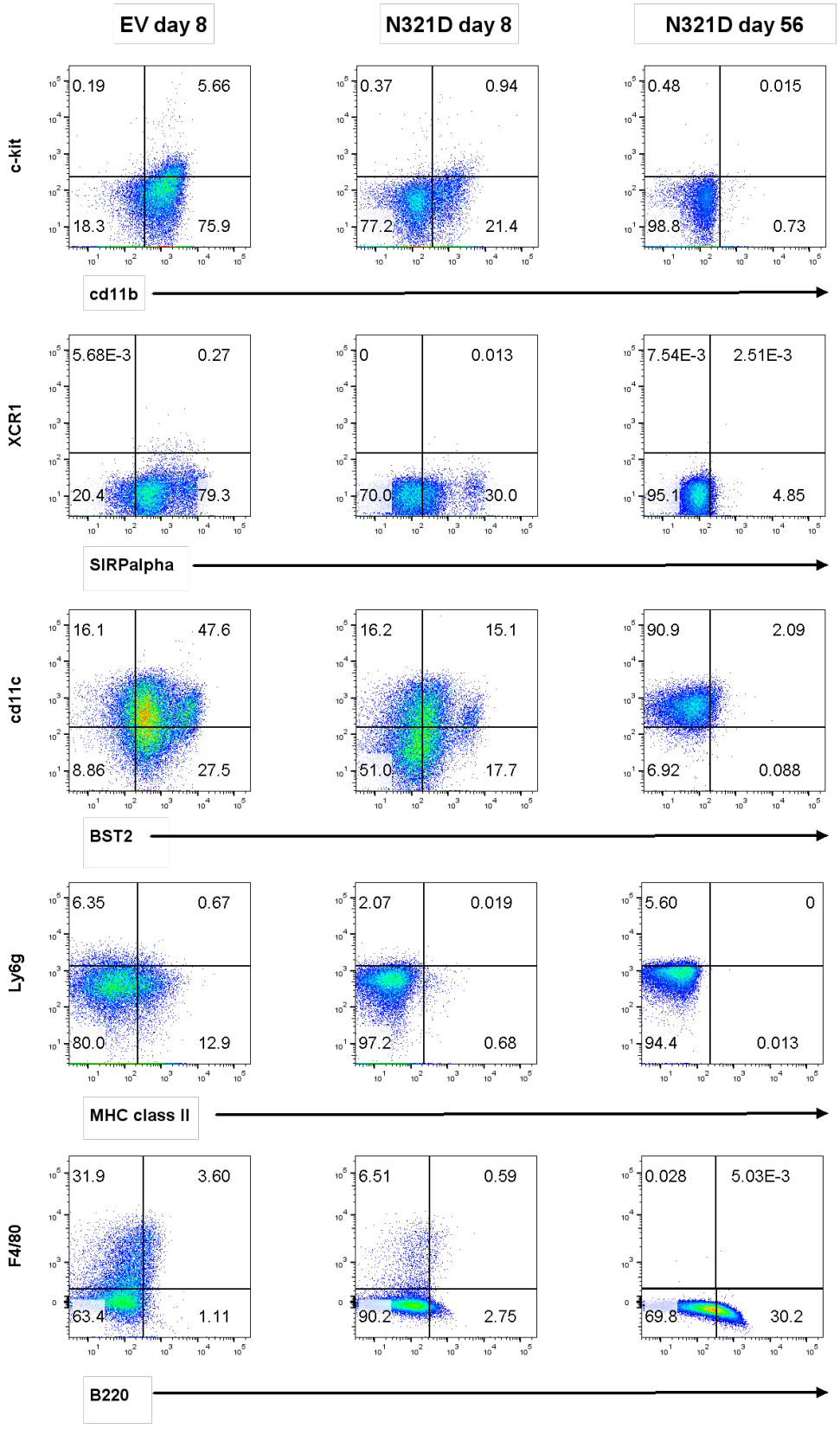


Figure 3.15 Differentiation profile of *CEBPA* N321D-transduced cells after extended time period in Hoxb8 media F+e- . *CEBPA* N321D-transduced cells are CD11c^{+} and B220^{+} at Day 56 of differentiation in Flt3L, in comparison with EV and mutant cells differentiated for 8 days. (1 replicate, 1 additional replicate of *CEBPA* N321D-transduced cells at Day 31 of differentiation with similar cell marker profile seen in Figure 3.17).

Intriguingly, the differentiation profile of the mutant cells has switched from reduced expression of cDC markers at week one, to a pDC progenitor-like phenotype at week eight. It was then decided to repeat differentiation of *CEBPA* N321D-transduced cells in Hoxb8 media F+e- and perform flow cytometry every 7 days from Day 3 to Day 31. This confirmed a gradual progression in the immunophenotype of mutant cells (see Figure 3.16).

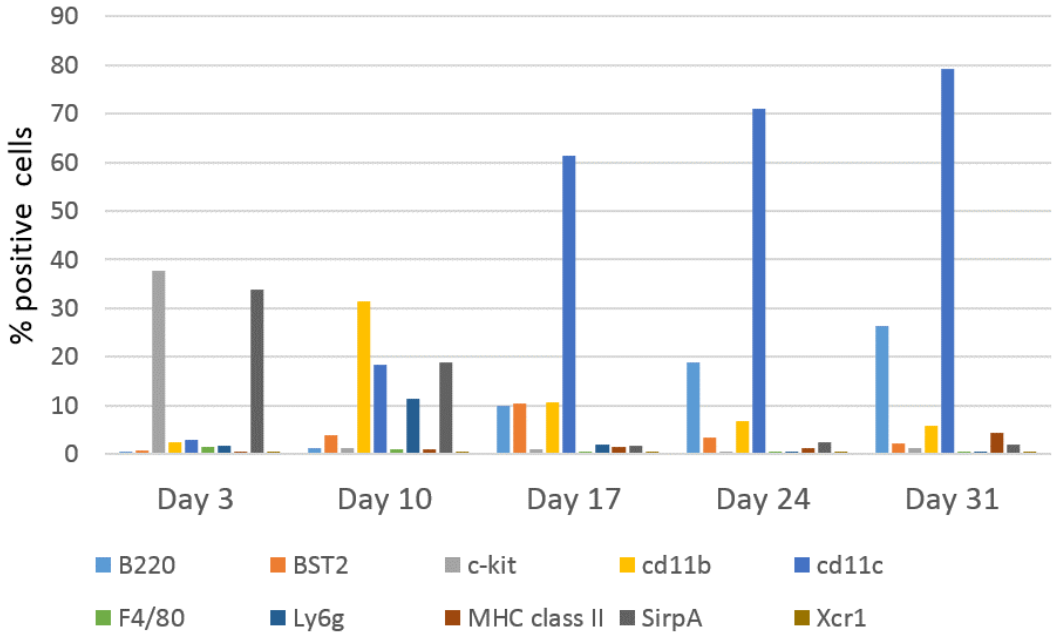


Figure 3.16 Flow cytometry shows progression of *CEBPA* N321D-transduced cell line through 4 weeks into a final c-kit ^{low} CD11c^{+} B220^{+} phenotype. (1 replicate only).

Flow cytometry at Day 31 with this biological replicate, originating from a distinct transfection of Hoxb8-FL cells with *CEBPA* N321D, demonstrated a similar profile as was previously observed (see Figure 3.15) at Day 56 with cells in Flt3L differentiating media:

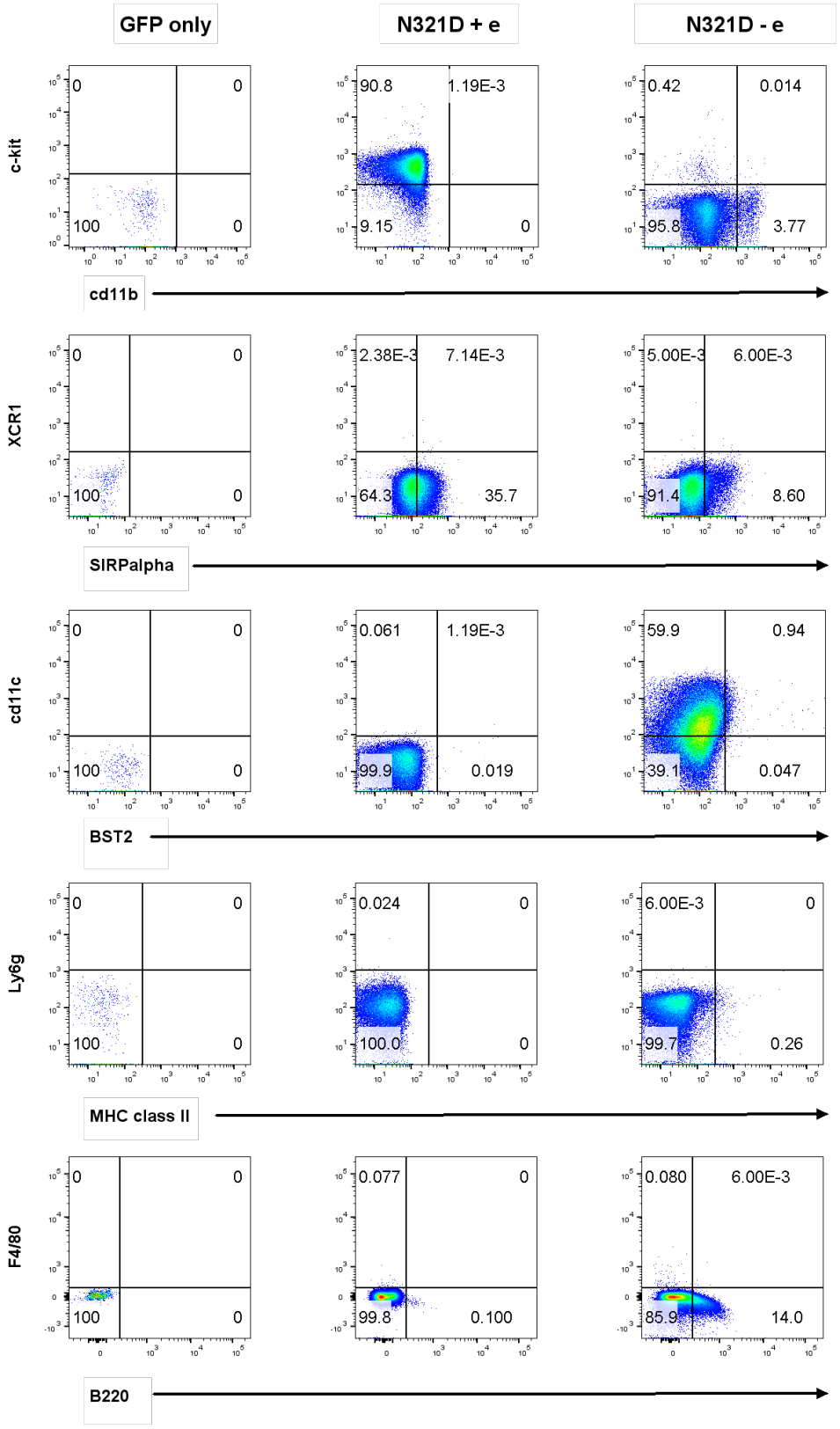


Figure 3.17 Differentiation profile of *CEBPA* N321D-transduced cells at Day 31 in Hoxb8 media F+e- . *CEBPA* N321D-transduced cells are confirmed as CD11c^{+} and B220^{+} at Day 31 of differentiation in Flt3L, using unstained GFP Hoxb8 parental cells in Flt3L and estradiol as a control. (1 replicate, 1 additional replicate of *CEBPA* N321D-transduced cells at Day 56 of differentiation with similar cell marker profile seen in Figure 3.15).

3.5 *In vivo* potential of Hoxb8-independent *CEBPA* N321D cells

Our *in vitro* results suggested that *CEBPA* N321D-transduced Hoxb8 cells were immortalised as a pDC progenitor-like cell line in the presence of Flt3L and in the absence of estradiol-mediated expression of *HOXB8*. We then sought to recapitulate these findings in an *in vivo* model. Hoxb8-FL cells were originally generated by Redecke *et al* from C57BL/6J mice and were able to transiently repopulate lethally irradiated mice in a time-dependent manner but did not demonstrate self-renewal capacity, leaving the bone marrow by Day 38. In addition, in a separate experimental model, bone marrow mononuclear cells were infected with retrovirus harbouring the *CEBPA* N321D mutation and induced a rapid leukaemic phenotype when transplanted into C57BL/6J mice (Togami *et al*, 2014).

We therefore decided to introduce Empty Vector and immortalised *CEBPA* N321D-transduced Hoxb8 cells into two cohorts of an *in vivo* study with five C57BL/6J mice in each cohort. EV-transduced cells in F+e+ media and *CEBPA* N321D-transduced cells at Day 38 of differentiation in Hoxb8 media F+e- were confirmed negative for mycoplasma before transplantation. Six-week old Ly 5.2 C57BL/6J mice were lethally irradiated and then injected with 5 x 10^5 Hoxb8-FL Ly 5.1 GFP^+ cells and 2 x 10^5 Ly 5.2 C57BL/6J bone marrow 'helper' cells via tail vein. Peripheral blood was withdrawn from the tail vein, initially weekly, to monitor engraftment and repopulation.

By Day 7 post-injection, the Ly 5.1 GFP^+ EV cells were undetectable in peripheral blood whereas the *CEBPA* N321D-transduced cells comprised about 9% of peripheral blood. However, the percentage of *CEBPA* N321D-transduced cells decreased steadily so that by Day 28 the mutant cohort showed <1.0% Ly 5.1 circulating cells (see Figure 3.18).

Mice from both cohorts demonstrated no Ly 5.1 cells in peripheral blood at Day 120, and exhibited no altered body weight, physical appearance or measurable clinical signs. It was therefore decided to sacrifice one mouse from each cohort at Day 149 post-infection. The mouse injected with *CEBPA* N321D-transduced cells demonstrated a distinct population of 6.2% $CD45.1^+$ $CD45.2^-$ cells in the bone marrow (Figure 3.19).

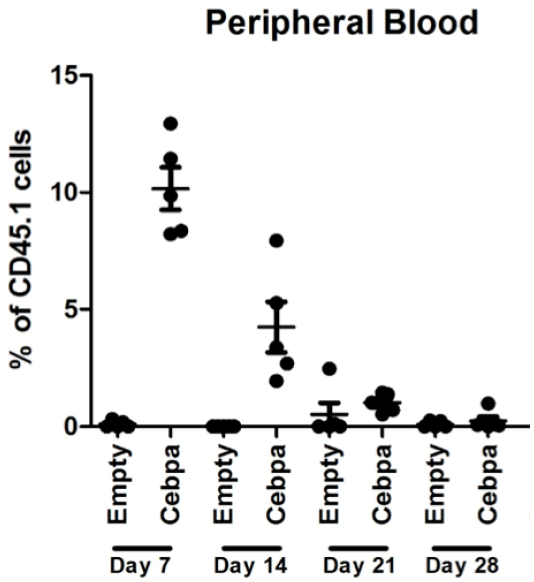


Figure 3.18 Circulating Ly 5.1 cells in C57BL/6J mice. Percentage of CD 45.1 cells in both cohorts of mice injected with EV and *CEBPA* N321D-transduced Hoxb8-FL cells. Figure provided by G Giotopoulos (Huntly group).

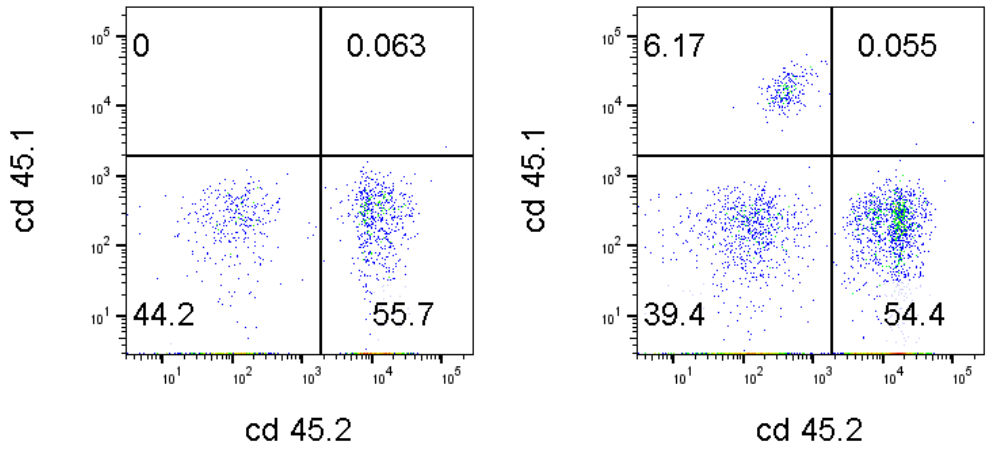


Figure 3.19 Bone marrow profile of C57BL/6J mice infected with CD 45.1 cells. (A) Mouse infected with EV-transduced Hoxb8-FL cell shows no long-term engraftment. (B) Mouse infected with *CEBPA* N321D-transduced cells demonstrates 6.2% engraftment of bone marrow cells at Day 149.

On Day 264, one of the mice reached humane end-point (weight loss, lethargy, hunched posture) and was culled. Peripheral blood showed mild anaemia and thrombocytopenia but

white cell count was normal (HGB 10.7 x 10^9/L, WBC 5.8 x 10^9/L, platelets 100 x 10^9/L). Necropsy revealed a marginally enlarged spleen (0.386 gr) and a normal sized liver (1.48 gr). In order to isolate leucocytes, particularly in the spleen, we gated on $CD45^+$ cells and found that these cells were 80.2% and 99.1% GFP^+ respectively in the bone marrow and spleen (Figure 3.20). Further characterisation of these GFP^+ cells demonstrated that they were B220 ^{high}, CD4 ^{low}, CD11b ^{high} and Gr1 ^{high} (Figure 3.21).

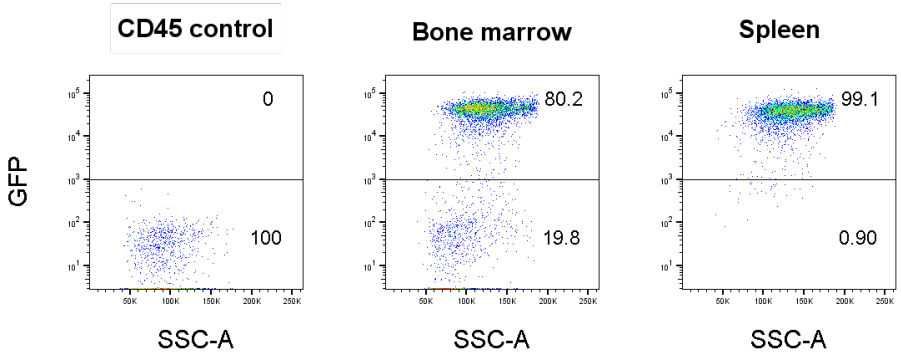


Figure 3.20 GFP^+ cells in bone marrow and spleen of mouse culled at Day 264. Cells were gated for $CD45^+$ leucocyte compartment, and demonstrated >20% infiltration of bone marrow and spleen with GFP^+ cells.

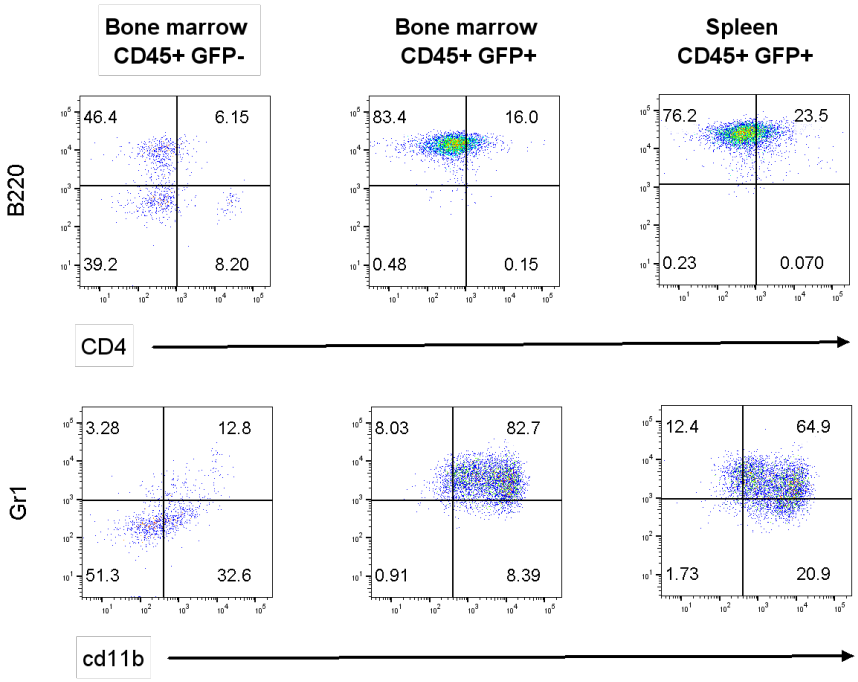


Figure 3.21 *CEBPA* N321D-transduced cells persist in bone marrow and spleen of mouse culled at Day 264 and express B220, CD11b and Gr1. These markers are consistent with DC phenotype.

This was a small preliminary study of two cohorts each of five mice infected with EV and *CEBPA* N321D-transduced cells respectively. None of the mice infected with EV cells showed evidence of engraftment during 12 months follow-up. In addition, none of the mice infected with *CEBPA* N321D-transduced cells demonstrated GFP^+ cells in peripheral blood at 4 months. However, one mouse was sacrificed from each arm of the study at 5 months post-infection, and the mouse infected with mutant cells showed evidence of long-term engraftment with 6.2% $CD45.1^+$ cells in the bone marrow, in comparison with no $CD45.1^+$ cells in the mouse infected with EV cells. As previously described, at 9 months one of the mice infected with *CEBPA* N321D-transduced cells demonstrated 80.2% and 99.1% infiltration of the $CD45^+$ leucocyte compartment in bone marrow and spleen respectively (Figure 3.20), showing markers consistent with DC phenotype (Figure 3.21). A third mouse infected with mutant cells died at month 11, and was found to have evidence of leukaemic infiltration of the bone marrow (data not shown). The two remaining mice infected with *CEBPA* N321D-transduced cells were culled at month 13: one showed expansion of GFP^+ cells >20% suggestive of leukaemia, but the other showed no evidence of long-term engraftment. Overall, these observations suggest that the *CEBPA* N321D-transduced cells are capable of long-term engraftment *in vivo*, but do not cause frank leukaemia on peripheral blood counts or significant organomegaly during 12 months of follow-up. In addition, the length of time required for leukaemic infiltration of the bone marrow suggests that additional hits by co-operating mutations may be required for a leukaemic phenotype to emerge.

3.6 Discussion

3.6.1 Cellular model of *CEBPA* expression

Our retroviral system is designed to answer the following question: What are the early changes associated with *CEBPA*-mutated pre-leukaemia? We have aimed to improve on previous cellular models of mutated *CEBPA* where the starting point has not been well-characterised, and consequently it has been difficult to compare start and end points. The Hoxb8-FL cell line was selected because it has significant advantages when building a cellular model of pre-leukaemia. Firstly, it is a conditionally immortalised cell line, with a fusion construct between EHRBD and Hox8 which provides a constant supply of near identical haematopoietic progenitor cells at a permissive temperature. Secondly, it is a robust cell line which is simple to both generate and maintain in cell culture. Thirdly, it comprises the

functional equivalent of LMPP-like cells, with the potential to differentiate into myeloid, lymphoid and dendritic cells *in vitro*. This contrasts with previous myeloid-committed progenitor cell lines employed by the Göttgens group such as the 416b cell line (Schutte *et al*, 2016) and the HPC7 cell line (Wilson *et al*, 2010).

The retention of identical mutation patterns in patients who relapse with *CEBPA*-mutated AML (Shih *et al*, 2006) suggests that these mutations are early events, and therefore it is useful to employ a cellular model which lies upstream of the CMP to GMP transition point. In addition, Naik *et al* have tracked developmental fates of early progenitor cells at single cell resolution by lentiviral-mediated barcoding, and identified not only lineage restriction in most LMPP cells, but also imprinting of B-cell, myeloid cell and DC lineages (Naik *et al*, 2013), suggesting that cell fate decisions occur in early haematopoiesis. Our hypothesis is that Hoxb8-FL cells can recapitulate the pre-leukaemic phenotype of early haematopoietic progenitors in a cellular model of mutated *CEBPA*.

The requirement for *CEBPA* expression at the CMP to GMP junction is well-characterised, with *CEBPA* deletion resulting in absolute loss of GMPs (Zhang *et al*, 1997) and macrophages (Heath *et al*, 2004). In addition, the Tenen group have demonstrated that *CEBPA* deletion causes severe depletion of dendritic cells when introduced into the LSK or GMP fraction, but has no effect when introduced into common dendritic progenitor cells (Welner *et al*, 2013). Interestingly, we find no significant distinction between the EV, *CEBPA* WT or *CEBPA* N321D-transduced cell lines on differentiation of early LMPP-like cells in either GM-CSF or IL-3, suggesting that GM-CSF receptor signalling may not be implicated in the pre-leukaemic stages of *CEBPA* N321D-mutated AML. Similarly, the lack of a distinct CD11b-related phenotype for *CEBPA* N321D-transduced cells in M-CSF implies that macrophage pathways are not disturbed by the N321D mutation at this stage of haematopoiesis. By contrast, the marked variation in phenotypic profile on differentiation of transduced Hoxb8-FL cell lines in Flt3L implicates a novel dendritic cell-mediated mechanism as an early effect of the *CEBPA* N321D mutation. It is important to bear in mind that this is one of many C-terminal mutations of *CEBPA* which are implicated in AML, that these mutations can affect functionally distinct leucine zipper or DNA-binding structures and that N-terminal mutations are likely to have very different effects on haematopoieisis. Moreover, the cell of origin is likely to have significant impact on leukaemic

phenotype, suggesting that expression of *CEBPA* N321D in progenitors distinct from the Hoxb8 model used here, may generate a different phenotype.

3.6.2 *CEBPA* N321D mutation

The *CEBPA* N321D mutation is found in human patients with AML (Papaemmanuil *et al*, 2016) and has previously been shown to induce rapid onset of aggressive leukaemia when lentivirally-transduced into BM mononuclear cells and injected into a mouse model (Togami *et al*, 2015). This C-terminal mutation gives rise to the p42 isoform of C/EBPα, preserving both transactivation domains and the DNA-binding domain. Single nucleotide substitution by N321D results in the in-frame mutation of asparagine to aspartate within the leucine zipper (LZ) domain of C/EBPα. The LZ is a three-dimensional structural motif formed from a heptad repeat $[abcdefg]_n$ where hydrophobic residues fall at positions *a* and *d* (typically leucine residues occur at position *d*), and polar/charged residues fall at positions *e* and *g*. A single residue can affect the configuration of the coiled coil structure of the leucine zipper, for instance the N321D mutation encodes aspartate in place of asparagine at the 321st codon, and asparagine residues prefer to pair at *a*-*a'* positions which can in turn reconfigure helix orientation (Grigoryan *et al*, 2008). The hydrophobic and polar interactions between these residues govern the helical topology of bZIP transcription factors and indeed engineered mutations within the LZ region of the MafB transcription factor have been shown to alter dimerisation properties and DNA-binding profile (Pogenberg *et al*, 2014). Given that the DNA-binding region of bZIP transcription factors is generally conserved and that combinatorial versatility is provided by the LZ region, it is likely that the N321D mutation confers altered selectivity for homo- and/or hetero-dimerisation profiles of C/EBPα, and that this in turn dysregulates *CEBPA*-mediated regulatory networks.

3.6.3 Novel pre-leukaemic dendritic cell phenotype induced by *CEBPA* N321D mutation

Dendritic cells (DCs) are exquisitely sensitive mononuclear phagocytes which bridge the innate and adaptive immune systems, responding to Toll-like receptor signalling, capturing proteins and presenting their peptide antigens to lymphocytes via MHC class I and MHC class II complexes. Conventional dendritic cells include type 1 cDCs which are $CD8^+$ $IRF8^+$ $XCR1^+$ and which specialise in cross-presentation of exogenous antigen to $CD8^+$ cytotoxic T cells, in addition to type 2 cDCs which are $CD8^-$ $SIRP\alpha^+$ and which stimulate $CD4^+$ helper T cells (Gurka *et al*, 2015). Plasmacytoid dendritic cells regulate T helper cells, generate T regulatory cells,

and secrete type 1 interferon on viral stimulation. Single-cell RNA-seq on human dendritic cells has recently identified novel DC subtypes and resolved functional inconsistencies by isolating a 'pure' pDC population (Villani *et al*, 2017).

The differentiation pathways of dendritic cells have proven challenging to decipher, with early work suggesting that cDCs and pDCs can develop from both myeloid and lymphoid-committed progenitors (Karsunky *et al*, 2003; Shigematsu *et al*, 2004; Ishikawa *et al*, 2007). The expression of Flt3 receptor for Flt3 Ligand is required for DC formation (Karsunky *et al,* 2003) but transcription factors such as IKAROS (Allman *et al*, 2006), GFI1 (Rathinam *et al*, 2005), and PU.1 (Carotta *et al*, 2010) also play essential roles. More recently, Schonheit *et al* showed that PU.1 remodels chromatin structure at the *IRF8* gene to initiate DC commitment, Lee *et al* showed lineage bias dependent on IRF8/PU.1 ratios in HSCs and LMPPs, and separately Welner *et al* showed that C/EBPα is required for early myeloid DC differentiation (Schonheit *et al*, 2013, Lee *et al*, 2017, Welner *et al*, 2013). IRF8, PU.1 and C/EBPα cross-talk informs cell fate decisions between neutrophil and monocyte lineages including DC lineages (Kurotaki *et al*, 2014). Core binding factors CBFβ and RUNX1 are required for the development of all $Flt3^+$ DC progenitors (Satpathy *et al*, 2014), whereas RUNX2 is required for terminal differentiation of pDCs (Sawai *et al*, 2013, Chopin *et al*, 2016). In addition, the Immunological Genome (ImmGen) Project has generated a compendium of microarray data which has allowed the identification of core gene signatures for cDCs, pDCs and other DC subsets (Miller *et al*, 2012).

It has been previously shown that Hoxb8-FL cells in Hoxb8 media F+e- differentiate by Day 6 into biphenotypic populations of conventional (defined as $CD11b^+$ $CD11c^+$ MHC class II^+ $B220^-$ $BST2^-$) and plasmacytoid ($CD11b^-$ $CD11c^+$ $B220^+$ $BST2^+$) dendritic cells. Our results do not replicate these findings exactly, in particular we do not identify $B220^+$ cells within 10 days of differentiation of the EV-transduced cells. This may reflect biological variability as Flt3L media was derived from pooled supernatant collected after seeding of B16-FLT3 melanoma cells, rather than from recombinant Flt3L. However, we do detect all other markers of cDCs and pDCs on differentiation of EV-transduced Hoxb8-FL cells, and our experiments identify that the cDCs are $SIRP\alpha^+$ type 2 cDCs rather than $XCR1^+$ type 1 cDCs.

CEBPA WT-transduced cells immediately promote a dendritic cell phenotype in Flt3L even in the presence of estradiol, suggesting that exogenous overexpression of *CEBPA* in Hoxb8-FL

cells is able to override the differentiation arrest caused by estradiol-mediated expression of *HOXB8*. Our results suggest that this phenotype is restricted to a subpopulation of *CEBPA* WT-transduced cells which express more GFP and therefore may be transduced with more plasmid copies, and that these GFP ^{high} cells differentiate, leaving behind cells transduced with lower numbers of *CEBPA* WT plasmid copies which persist in estradiol and behave phenotypically like Hoxb8 parental cells. EV and *CEBPA* N321D-transduced cells remain c-Kit^{+} and fail to express other markers of myeloid, lymphoid or dendritic cell maturation in F+e+ media. Both the *CEBPA* WT and N321D-transduced cells demonstrate a growth disadvantage in competitive assay with Hoxb8-FL parental cells over time, but persist at low titres long-term.

On withdrawal of estradiol, both EV and *CEBPA* WT-transduced cells rapidly differentiate into cDCs and pDCs, and eventually are undetectable by Casy Cell Counter. By contrast, *CEBPA* N321D-transduced cells demonstrate maturation arrest within 5 to 10 days, which is characterised by markers of cDC (CD11b, SIRPα) but also pDC (CD11c, BST2) lineage specification. Importantly, the *CEBPA* N321D-transduced cells appear to become immortalised in a Hoxb8-independent fashion, and gradually undergo phenotypic transformation with a CD11c^{+} B220^{+} pDC-like profile on flow cytometry by Day 24 of differentiation. These immortalised cells are BST2^{-}, which may reflect a pDC progenitor status. Interestingly, our result of an initial cDC phenotype progressing to a late pDC progenitor phenotype reverses observations made in early studies which suggested that pDCs could differentiate into cDCs *in vitro* (Grouard *et al*, 1997). Grouard *et al*'s observations have subsequently been interpreted as short-lived pDCs declining and differentiating cDCs coming to dominate in cell culture, whereas our observations reflect the dominance of an immortalised pDC-like progenitor and the gradual decline of differentiating cDCs. An important question not addressed by our experimental work is how this immortalised pDC progenitor may contribute mechanistically to the myeloid leukaemic phenotype seen in AML. In particular, mouse models have shown that IFN-α is essential for T-cell priming and tumour elimination via antigen cross-presentation to CD8^{+} T-cells (Diamond *et al*, 2011) and several clinical trials have shown evidence that IFN-α has anti-leukaemic effects (Smits *et al*, 2013; Mo *et al*, 2018). In addition, there is emerging evidence of pDC and cDC cross-talk in cancer immunity (Noubade *et al*, 2019) and there is the possibility that our pDC progenitor phenotype deregulates cancer antigen presentation by conventional dendritic cells. Future experiments may elucidate the pDC functional status of *CEBPA* N321D-transduced cells, for instance it

would be interesting to verify whether CpG-mediated stimulation of immortalised pDC progenitor cells promotes secretion of IFN-α or if co-culture with $CD4^+$ T-cells promotes T-cell proliferation.

On transplantation, the *CEBPA* N321D-transduced cells are capable of long-term engraftment in an *in vivo* C57BL/6J mouse model. However, initially on sacrificing a mouse at 5 months, there are no circulating cells in peripheral blood, blood counts are normal and there is no splenomegaly. Long-term follow-up of transplanted mice shows increased Ly 5.1 cells consistent with leukaemia in the bone marrow and splenic compartments at a relatively late stage in three out of five mice, but little effect on peripheral blood counts or organomegaly, suggesting that additional cooperative mutational hits may be required for leukaemic phenotype. Interestingly, this varies from the leukaemic phenotype demonstrated on transplantation of C57BL/6J mice with cells infected with retroviruses harbouring the *CEBPA* N321D mutation by Togami *et al*. However, there is evidence that C/EBPα influences cell fate decisions in a highly state dependent fashion (Zhang *et al*, 2004) and it should be remembered that Togami *et al* transduced a heterogeneous population of BM mononuclear cells, whereas our experiment infects LMPP-like cells. In addition, Togami *et al* employed a different retroviral model which may well have had different transduction efficiency resulting in different expression levels compared to our approach. By contrast, the Nerlov group found that a knock-in model of bi-allelic C-terminal mutations expanded the HSC pool and skewed lineage programming of HSCs but did not cause leukaemia (Bereshchenko *et al*, 2009). It seems reasonable to speculate that *CEBPA* N321D-mediated maturation arrest in the dendritic cell compartment may represent a pre-leukaemic initiating event which requires secondary mutations for the development of leukaemia in a multi-step process. Double transgenic mouse models of *CEBPA* N321D with either a N-terminal mutation of *CEBPA* or alternatively with mutations known to commonly cooperate with single-mutation *CEBPA* in human AML (Fasan *et al*, 2014) may validate this hypothesis, and it would also be interesting to perform secondary transplants of leukaemic cells *in vivo* to see if a more proliferative leukaemic effect is generated in secondary recipients, but both these potential projects are beyond the scope of this study. To our knowledge, this is the first time that a dendritic cell mechanism has been implicated as a pre-leukaemic change in mutated *CEBPA*.

4. Single cell gene expression analysis reveals heterogeneity in *CEBPA* N321D-transduced cells and dysregulation of HSPC and DC transcriptional programs

4.1 Background

4.1.1 Global gene expression in a cellular model of mutated *CEBPA*

Our hypothesis is that a mutated version of *CEBPA* will perturb C/EBPα-dependent regulatory networks and that the changes can be characterised by performing single cell RNA-seq. For this purpose, we used LMPP-like cells transduced with retroviruses harbouring Empty Vector (EV), *CEBPA* WT and *CEBPA* N321D. Initial characterisation by microscopy and flow cytometry identified that differentiation in the presence of Flt3L demonstrated the most significant phenotypic effects, and therefore this condition merits more detailed exploration at the level of the single cell.

Transcriptional regulation plays an intrinsic role in the cell fate decisions which determine the differentiation of self-renewing multipotent HSCs through progressive stages of lineage restriction into distinct subsets of mature blood cells. The potency of this transcriptional machinery has been demonstrated by TF-mediated reprogramming of differentiation pathways from mature somatic cells to pluripotent stem cells (Takahashi *et al*, 2006).

RNA-seq has become ubiquitous in transcriptomics since it was first described in 2008 (Nagalakshmi *et al*, 2008). It has a number of advantages over previous microarray technology including a higher signal-to-noise ratio, a larger range of expression values, and the ability to map gene expression against the entire genome. RNA-seq can be combined with fluorescence activated cell sorting (FACS) to describe transcription profiles of distinct populations within the haematopoietic tree.

RNA-seq workflows are designed to favour either annotation of novel transcriptomic data or alternatively differential gene expression (DGE). It is important to identify a hypothesis before designing an experimental workflow, for instance DGE workflows prioritise sequencing depth (Tarazona *et al*, 2011) and biological replicates (Hansen *et al*, 2011). DGE experiments aim to quantify differences in transcript abundance between multiple conditions or groups, and

therefore have to account for variances within complex biological systems. Even isogenic cells are subject to biological variance from transcriptional burst kinetics due to factors such as chromatin environment (Dey *et al*, 2015), DNA looping (Bartman *et al*, 2016), histone modifications (Wu *et al*, 2017) and the affinity of regulatory elements (Suter *et al*, 2011). Biological replicates can account for variance and are essential for the interpretation of RNA-seq data on differential expression.

RNA-seq experimental protocols commence with disruption of the cell membrane, followed by RNA extraction using either organic solvents or solid-phase extraction. Targeted enrichment of mRNA is achieved typically by hybridisation of poly-adenyl tails to oligo-deoxythymines (oligo-dTs). RNA fragments are then used as templates for reverse transcriptase-mediated cDNA synthesis. The ends of the cDNA fragments are then ligated with sequencing adapters, which can include indexed barcodes for pooling in multiplex libraries. The cDNA libraries are quantified before sequencing, and data is then run through an RNA-seq analysis pipeline which typically includes filtering, mapping to a reference genome, normalisation, multivariate statistical analysis, and data visualisation.

4.1.2 Single-cell analysis

Single-cell technology presents significant advantages over previous genomic tools which analysed bulk tissue samples and inferred cell types and trajectories from averaged datasets derived from millions of cells.

Firstly, single-cell technology permits transcriptional analysis in the presence of paucity of cells, particularly during early developmental biology. For instance, Scialdone *et al* studied 1,205 single cells to identify early lineage diversification points at the epithelial-to-mesenchymal transition (Scialdone *et al*, 2016).

Secondly, evaluating transcription profiles at the level of the single cell can dramatically increase our understanding of differentiation pathways and their deregulation, based on the premise that cell fate decisions are essentially unicellular. This premise was originally demonstrated by transplantation of single HSCs (Osawa *et al*, 1996). Retroviral integration of reporter genes (Turner *et al*, 1987) and cell-specific viral DNA barcoding (Lu *et al*, 2011) subsequently permitted vector linear tracing of single cells. Examples of novel differentiation

pathways elucidated by single-cell RNA-sequencing (scRNA-seq) include the early diversification between mast cell and basophil progenitors (Dahlin *et al*, 2018). Cancer stem cells are single malignant clones (Quintana *et al*, 2008), and therefore intuitively single-cell technology presents significant opportunities for understanding oncogenesis. For instance, single cell technology has described a subset of CML stem cells which are selected out for survival during prolonged therapy (Giustacchini *et al*, 2017).

Thirdly, scRNA-seq provides a powerful tool to explore heterogeneity within cell populations that appear homogenous using cell surface markers. For instance, massively parallel scRNA-seq characterised 19 novel subpopulations amongst myeloid progenitors after index sorting by conventional surface markers (Paul *et al*, 2015). More recently this technology has been applied to human dendritic cells (DCs) to identify a novel subset of DCs and redefine plasmacytoid dendritic cells (pDCs) (Villani *et al*, 2017). This transcriptional heterogeneity can then be probed for functional relationships, for instance scRNA-seq was followed by functional studies to demonstrate that the first wave of IFN-α production depends on cues from the microenvironment and is limited to a subset of early responder pDCs (Wimmers *et al*, 2018). Haematological malignancies comprise particularly interesting candidates for single-cell analysis because intratumoral heterogeneity underpins fundamental mechanisms of cancer evasion. For instance, single leukaemic cells from AML patients with Quizartinib resistance demonstrate significant clonal diversity (Smith *et al*, 2017).

Single-cell technology is rapidly evolving with a range of scRNA-seq methods such as Smart-seq2 (Picelli *et al*, 2013), Drop-seq (Macosko *et al*, 2015) and commercial platforms such as 10x Genomics Chromium. Each technique has advantages and disadvantages including sensitivity, precision, multiplexing, the addition of unique molecular identifiers (UMIs), RNA processing, and cost-efficiency. Smart-seq2 is widely used for small numbers of single cells because it is highly sensitive, accurate, has lower dropout probability, and enables full-length capture of mRNA, although with significant 3' bias during cDNA generation by oligo-dT primers. Disadvantages include the difficulty in adding UMIs to account for amplification bias, and that the manual procedure requires careful experimental design to avoid batch effects.

Single-cell RNA-seq data analysis differs from bulk RNA-seq because each single cell is represented by a distinct sequencing library, and normalisation is particularly important to

deal with amplification bias and gene dropouts between libraries. ERCC spike-in are commonly employed in Smart-seq2 to improve normalisation. Differential expression and clustering can be carried out using methods originally developed for bulk RNA expression analysis such as DESeq2 (Love *et al*, 2014), but novel tools including SCDE (Kharchenko *et al*, 2014) and DESingle (Miao *et al*, 2018) have evolved specifically to analyse scRNA-seq data. Performance is equivalent for both bulk and single cell analytical tools (Soneson *et al*, 2018).

Data visualisation of single-cell transcriptomic data typically involves analysis by dimensionality reduction and clustering algorithms. Principal Component Analysis (PCA) resolves high-dimensional data onto lower-dimensional linear space (typically 2 or 3 dimensions) by establishing orthogonal projections along which the variance of data is high (Hotelling, 1933). The first principal component is the direction in space along which data points have the largest variance, the second principal component is the direction which maximises variance among all projected data points orthogonal to the first, and so on. The PCA analysis creates as many components as there are variables (in this case, genes with read counts) and each component is a linear combination of all variables, but normally only the first two or three principal components are used to visualise gene expression data because they explain most of the variability between cells. This statistical method can be employed to explicitly map and project complex genomic data onto lower dimensions for visualisation. However, PCA-based projections are restricted to linear transformations of high-dimensional data and therefore can fail to project datasets where data is not linearly correlated, for instance if data points lie upon a spiral-like manifold. Furthermore, the complex structure of gene expression data is often better captured in 2 or 3 dimensions by non-linear methods (Bartenhagen *et al*, 2010). In this regard, t-Distributed Stochastic Neighbour Embedding (t-SNE) converts high-dimensional data into a matrix of pairwise similarities between all cells, and visualises the similarity data by first specifying a limited number of components to represent the data, and then identifying positions for cells in these components that preserve these similarities between pairs of cells (van der Maaten *et al*, 2008). t-SNE employs Euclidean distances or other similarity measures to learn about discrepancies between pairs of data and is therefore better able to preserve local data architecture and cluster cells which are transcriptionally similar. For these reasons, it is unsurprising that transcriptomic relationships and variances within clusters at the level of the single cell are often better visualised using t-SNE rather than PCA. More recently, UMAP (McInnes *et al*, 2018) has been developed as a

data visualisation method that preserves global data architecture as well as preserving local neighbour relations, and SPRING generates k-nearest neighbour graphs to visualise data using force-directed layouts which are more reproducible than t-SNE and better able to visualise continuous expression topologies (Weinreb *et al*, 2018).

Lastly, scRNA-seq can be employed combinatorially with other technologies to further characterise gene expression at the level of the single cell. Methods such as Perturb-Seq (Dixit *et al*, 2016) and CRISPR-Seq (Tothova *et al*, 2017) combine single-cell transcriptome analysis with gene perturbations. Similarly, scRNA-seq can be combined with chromatin accessibility data to construct a chromatin accessibility landscape of haematopoiesis and establish novel differentiation trajectories (Buenrostro *et al*, 2018).

4.2 Experimental design

Hoxb8 parental cells were thawed and cultured for 7 days, and then transduced with retroviral supernatant to prepare three biological replicates each of EV, *CEBPA* WT and *CEBPA* N321D, as described in Chapter 3. Cells were then taken at Day 2 post-transduction for differentiation, washed twice in PBS, placed in wells at 1 x 10^5 cells/ml in 2 ml of Hoxb8 media with Flt3L and no estradiol (F+e-), and cultured for a further 5 days until Day 7 post-transduction. Single cell-sorting of 32 cells per replicate at two time points (Days 2 and 7 post-transduction, or Days 0 and 5 post-differentiation) was performed on a BD Influx cell sorter running BD FACS software, alongside flow cytometry with a basic panel of markers (c-kit, Ly6g, CD11b, and B220), and the 576 single cells were then processed as one batch using a modified Smart-seq2 protocol described by Picelli *et al* (Picelli *et al*, 2014). Replicates were evenly sorted into 6 plates and evenly distributed in 3 lanes on HISeq4000, aiming to reduce sources of technical variation by employing a balanced block design (Baran-Gale, 2018). Following scRNA-seq, transcripts were aligned to the mouse genome using GSNAP and Ensembl archive 81, with specific additional annotation for GFP and for human *CEBPA*. It should be noted that alignment to human *CEBPA* did not reliably distinguish the single point mutation between human *CEBPA* WT and human *CEBPA* N321D. Quality control was then applied with the following criteria: 3,000,000 > total reads > 200,000; fraction mapped > 0.15; highly expressed genes (defined as counts per million > 50) > 1800; ERCC fraction < 0.12; and mitochondrial gene fraction < 0.2.

Only the 479 cells that pass quality control were kept for subsequent analysis. The cells were distributed as follows:

Table 4.1 Quality Control for Cell type, biological replicate, and day/condition

Sample-Replicate-Day	Total Reads	Mapped Reads	Mapped Fraction	ERCC Fraction	Highly expressed genes	Mito-chondrial Fraction	QC passed cells
EV-R1-D2	63362912	33712501	0.5	0.03	3571.91	0.08	30
CebpaWT-R1-D2	49475555	20977838	0.41	0.04	2912.39	0.07	25
CebpaN321D-R1-D2	51183404	24131687	0.44	0.04	3018.58	0.1	25
EV-R2-D2	49320596	26084451	0.52	0.03	3306.53	0.09	29
CebpaWT-R2-D2	50558382	26811977	0.5	0.05	2893.85	0.08	29
CebpaN321D-R2-D2	66548088	37163956	0.55	0.02	3745.58	0.07	31
EV-R3-D2	68018455	37955758	0.54	0.02	3713.91	0.08	31
CebpaWT-R3-D2	41734500	20796897	0.43	0.06	2416.3	0.09	21
CebpaN321D-R3-D2	41423878	20898580	0.46	0.05	2827.26	0.1	24
EV-R1-D7	57308244	28164329	0.44	0.08	2241.38	0.11	24
CebpaWT-R1-D7	40481057	15880198	0.28	0.15	1480.82	0.11	12
CebpaN321D-R1-D7	60829038	29085480	0.42	0.07	2649.77	0.1	23
EV-R2-D7	43098006	21604380	0.48	0.06	2614.69	0.08	26
CebpaWT-R2-D7	66943615	37563159	0.55	0.03	3257.55	0.06	33
CebpaN321D-R2-D7	79005298	42265521	0.52	0.03	3220.39	0.07	30
EV-R3-D7	53470150	28480222	0.51	0.03	3182.84	0.06	28
CebpaWT-R3-D7	52445696	27880078	0.51	0.03	3092.15	0.04	30
CebpaN321D-R3-D7	56999382	31592893	0.52	0.04	3198.39	0.06	28

4.3 Flow cytometry of sorted cells confirms CD11b phenotype before scRNA-seq

Cells from each replicate were stained for c-kit, Ly6g, CD11b and B220, and then processed with BD FACS software. In line with the results described in Chapter 3, the only cell surface marker which showed a significant effect was CD11b, which was upregulated in all replicates of the *CEBPA* WT-transduced cells at Day 2 post-transduction and showed reduced expression in all replicates of the *CEBPA* N321D-transduced cells at Day 7:

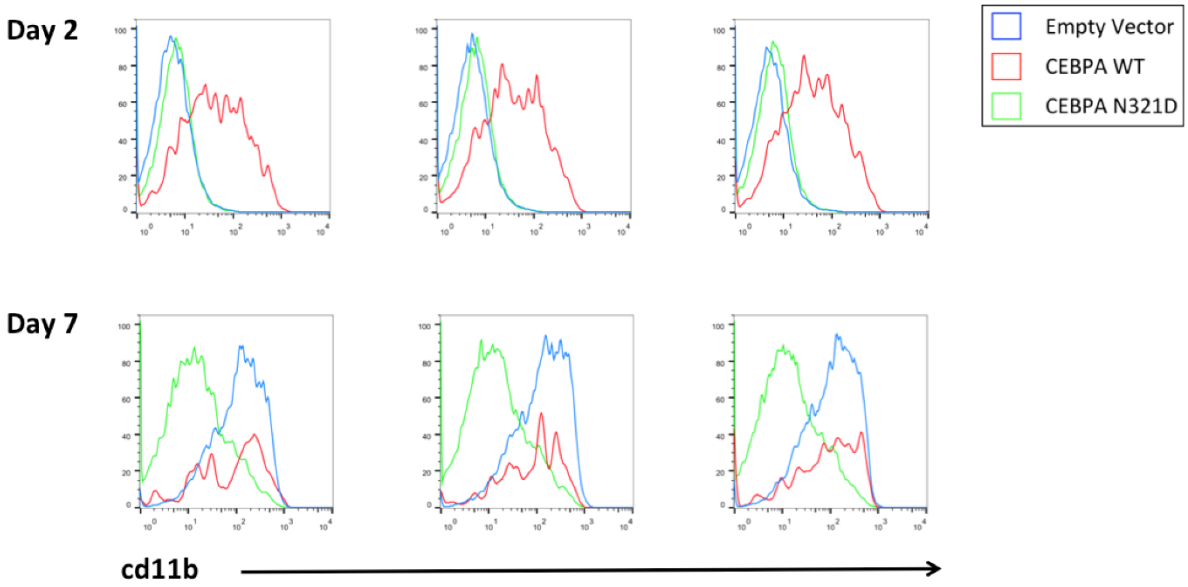


Figure 4.1 CD11b expression (normalized to mode) for 3 biological replicates of EV, *CEBPA* WT and *CEBPA* N321D-transduced cells at Days 2 and 7 post-transduction before index-sorting

4.4 Analysis of single-cell data using the Automated Single-cell Analysis Pipeline (ASAP)

ASAP is a web-based platform for the analysis and visualisation of single-cell RNA-seq data (Gardeux *et al*, 2017). It provides a visual web-based interface for a single-cell pipeline which allows the user to analyse data after genome alignment. It performs QC, provides different normalization methods, allows visualisation including PCA and t-SNE, resolves cells into clusters, and calculates differential expression between defined clusters. It also provides gene enrichment by Gene Ontology or KEGG. One important limitation is that it can only perform differential expression against the genome of a defined species, and is unable to process exogenous gene inserts such as GFP, but nevertheless it provides a rapid and flexible tool for non-bioinformaticians to analyse single-cell data.

4.4.1 Normalisation of single-cell data

The single-cell Latent Variable Model (scLVM) was used to normalise data and account for confounding factors such as cell cycle variation which can introduce noise into single-cell data sets (Buettner *et al*, 2015). Gene expression was evenly distributed across cell samples after normalisation (Figure 4.2). ERCC spike-ins were used for fitting of technical noise (Figure 4.3),

and the method of Brennecke *et al* (Brennecke *et al*, 2013) was employed to infer 10,456 highly variable genes (Figure 4.4).

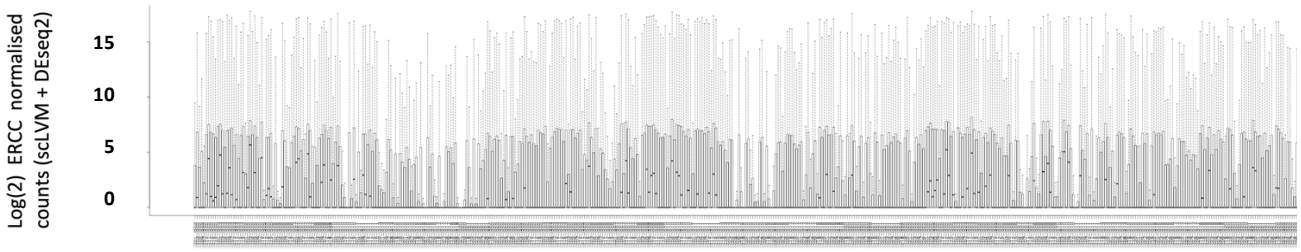


Figure 4.2 Distribution of the expression of genes in each cell after normalisation

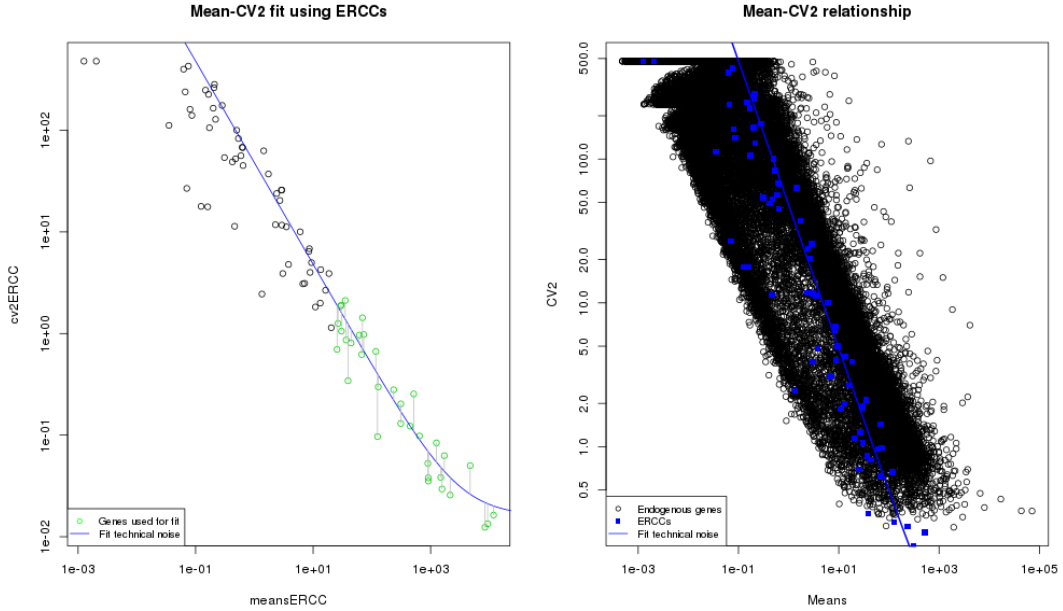


Figure 4.3 scLVM fit of technical noise

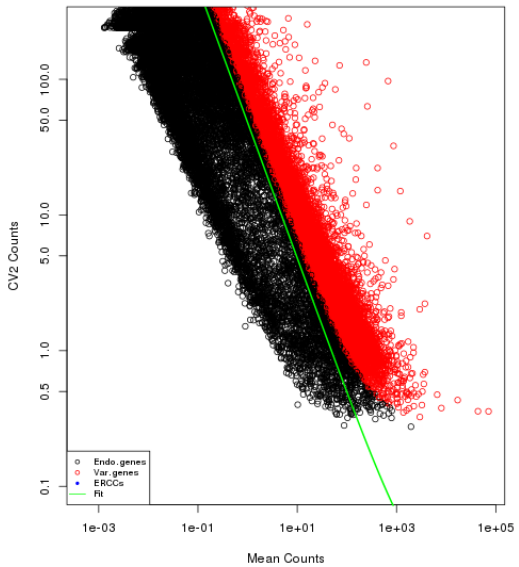


Figure 4.4 Variable genes identified after scLVM fit of technical noise

4.4.2 Single-cell data identifies distinct distributions of *CEBPA* WT and N321D-transduced cells and subpopulations within *CEBPA* N321D-transduced cells

Transcription profiles of single cells can be employed to cluster cells and annotate novel cell types by dimensionality reduction and unsupervised hierarchical clustering. ASAP facilitates PCA as a classical linear scaling method for dimensionality reduction, and t-SNE for non-linear dimensionality reduction (van der Maaten *et al*, 2008). Both techniques were employed, but it was found that t-SNE gave a more informative visual representation of distinct cell populations. Cell identities were manually defined, and t-SNE was performed with perplexity 25 and theta 0.5 to visualise the data both within ASAP's own internal visualisation function (Figure 4.5), and also by downloading the data for visualisation using the R-package ggplot2 (Figure 4.6).

Several observations can be made from visual representation of the single-cell data. Firstly, EV-transduced cells demonstrate a distinctive transcriptional signature in their non-differentiated state (corresponding phenotypically to Hoxb8-FL LMPP-like cells) as compared to after 5 days of differentiation in Flt3L, with no overlap between the two conditions. This conforms with the distinct cell surface marker phenotype with upregulated CD11b in differentiated EV cells described in Chapter 3.

Secondly, at Day 2 post-transduction, in conditions which normally maintain Hoxb8-FL cells in a non-differentiated state, it is clear that the *CEBPA* WT-transduced cells form a distinct cluster on t-SNE which is very different to EV cells maintained at Day 2 in Flt3L and estradiol. This correlates with flow cytometry data described in Chapter 3 where the *CEBPA* WT-transduced cells express CD11b and SIRPα even before abolishing estradiol-mediated nuclear translocation of Hoxb8, whereas the EV and *CEBPA* N321D-transduced cells express no lineage markers in the presence of Flt3L and estradiol. Therefore, expression of *CEBPA* WT pushes Hoxb8-FL cells towards a distinctive transcriptional state.

Thirdly, it becomes difficult to visually distinguish *CEBPA* WT-transduced cells from EV cells by t-SNE after 5 days of differentiation, which again correlates with flow cytometry data. This suggests that the overexpression of human WT *CEBPA* in Hoxb8-FL cells cannot by itself overcome endogenous mechanisms of Flt3L-mediated dendritic cell (DC) differentiation following withdrawal of estradiol.

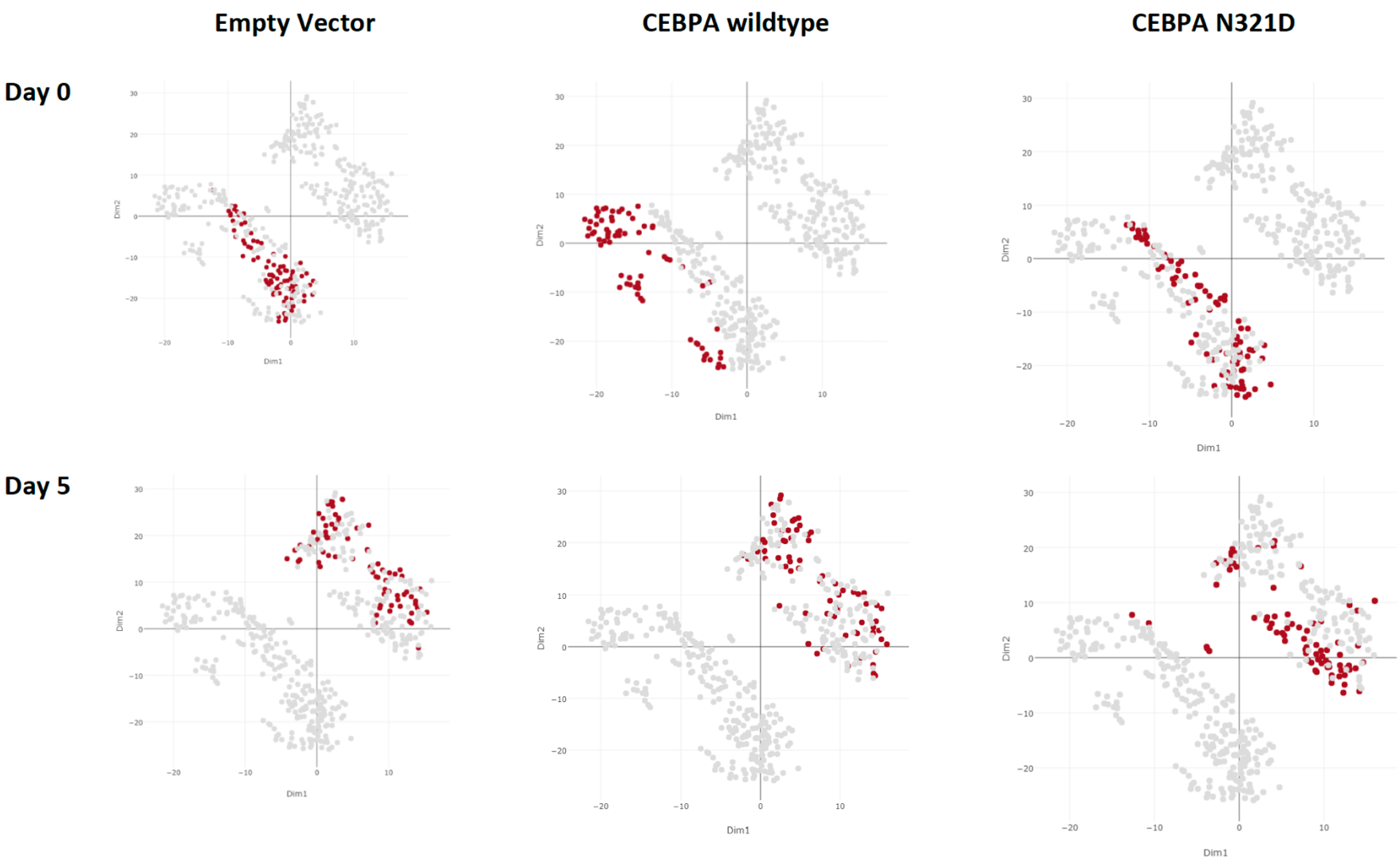


Figure 4.5 ASAP web-based t-SNE visual representation of EV, CEBPA WT and CEBPA N321D-transduced single cells at Days 0 and 5 of differentiation (equivalent to Days 2 and 7 post-transduction)

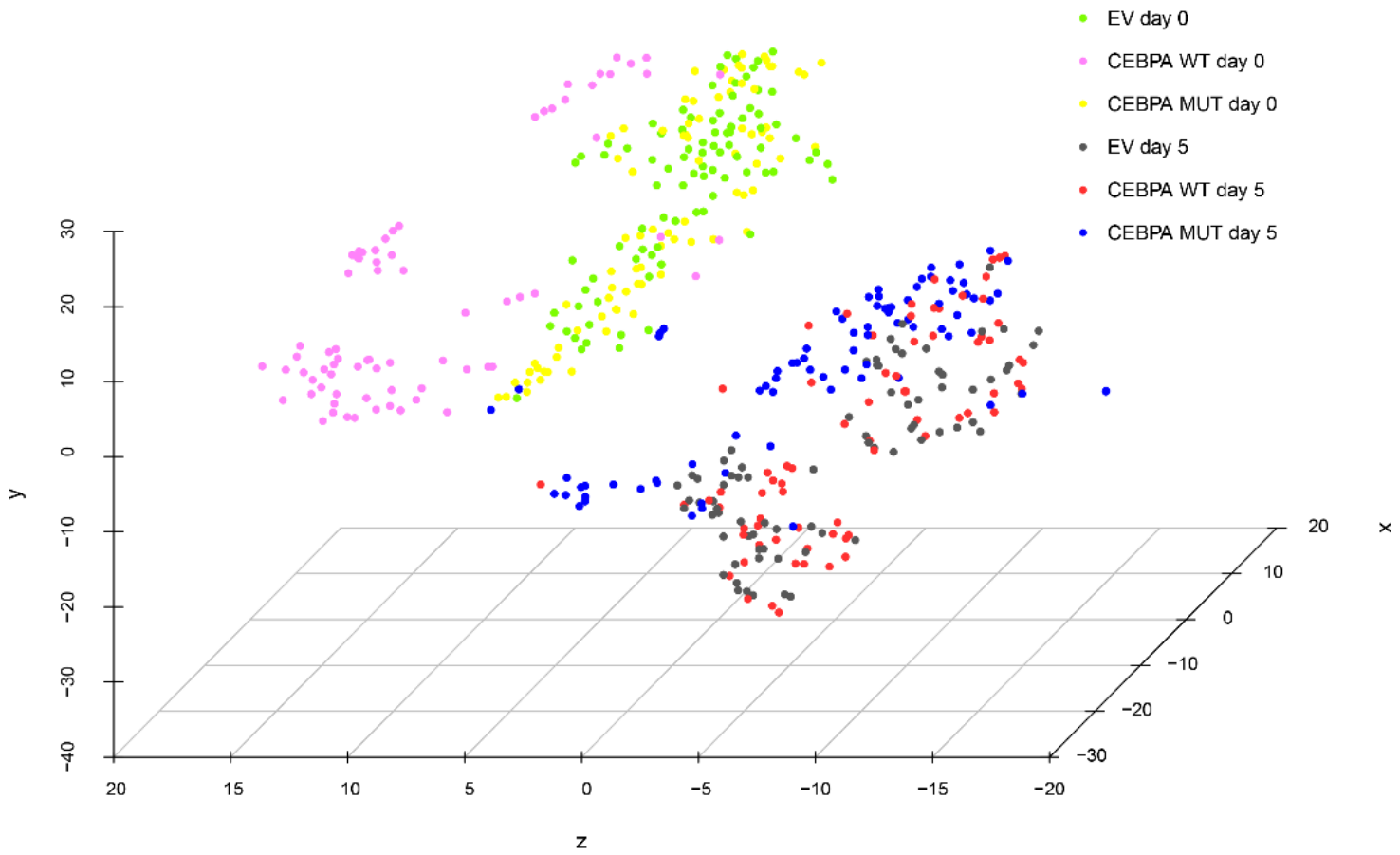


Figure 4.6 ASAP single cell data plotted using ggplot2 for 3D visual representation. Data shows cells at Days 0 and 5 of differentiation

Fourthly, *CEBPA* N321D-transduced cells demonstrate a similar transcriptional profile to EV cells in the presence of estradiol, suggesting that *CEBPA* N321D cannot push Hoxb8-FL cells towards the altered transcriptional profile reached by *CEBPA* WT-transduced cells in the presence of nuclear Hoxb8.

Fifthly, *CEBPA* N321D-transduced cells diverge from both EV and *CEBPA* WT-transduced cells after 5 days of differentiation in Flt3L, and again this corresponds to their very distinct expression profile on probing with fluorescent-labelled antibodies. Furthermore, following the withdrawal of estradiol, two distinct subpopulations can be identified within the *CEBPA* N321D-transduced cell population, with the majority of cells distributed proximally within the Day 5 differentiated cells, but a much smaller cluster of 5 *CEBPA* N321D-transduced single cells distributed amongst or very close to the Day 0 non-differentiated EV-transduced cells.

In summary, the exogenous expression of *CEBPA* WT pushes LMPP-like cells towards a distinct transcriptional state in the presence of nuclear Hoxb8 but does not interfere with endogenous mechanisms of differentiation when Hoxb8 translocates to the cytoplasm, whereas the exogenous expression of *CEBPA* N321D cannot disrupt Hoxb8 transcriptional effects but is able to alter the endogenous differentiation trajectory of LMPP-like cells in the presence of Flt3L.

4.4.3 The gene expression profiles of Hoxb8-FL cells differentiated in Flt3L correspond to ImmGen gene expression modules for pDCs and a specific cohort of cDCs

Phenotypic data from flow cytometry experiments described in Chapter 3 has demonstrated that EV-transduced Hoxb8-FL cells differentiated in Flt3L express antigens for CD11b and SIRPα but not XCR1 after 5 days of differentiation, suggesting a conventional dendritic cell (cDC) subset which have been variously described as CD8α^- or type 2 cDCs. However, *CEBPA* N321D-transduced cells demonstrate reduced expression of CD11b and SIRPα antigen from Day 5 of differentiation in Flt3L, and gradually take on a $CD11c^+$ $B220^+$ $BST2^-$ pDC precursor-like phenotype. scRNA-seq presents the opportunity to correlate this immunophenotype with gene expression data.

The ImmGen Consortium has employed a multicentre strategy to build a comprehensive gene expression database of murine haematopoiesis, and has defined 334 fine modules of genes

with significant co-expression (http://www.immgen.org/ModsRegs/modules.html). The modules F150, F156 and F152 show substantial upregulation in pDCs, cDCs and CD8α^+ CD103$^+$ cDCs respectively (Miller *et al*, 2012), and expression profiles for key genes in these modules can be plotted within our single-cell data set:

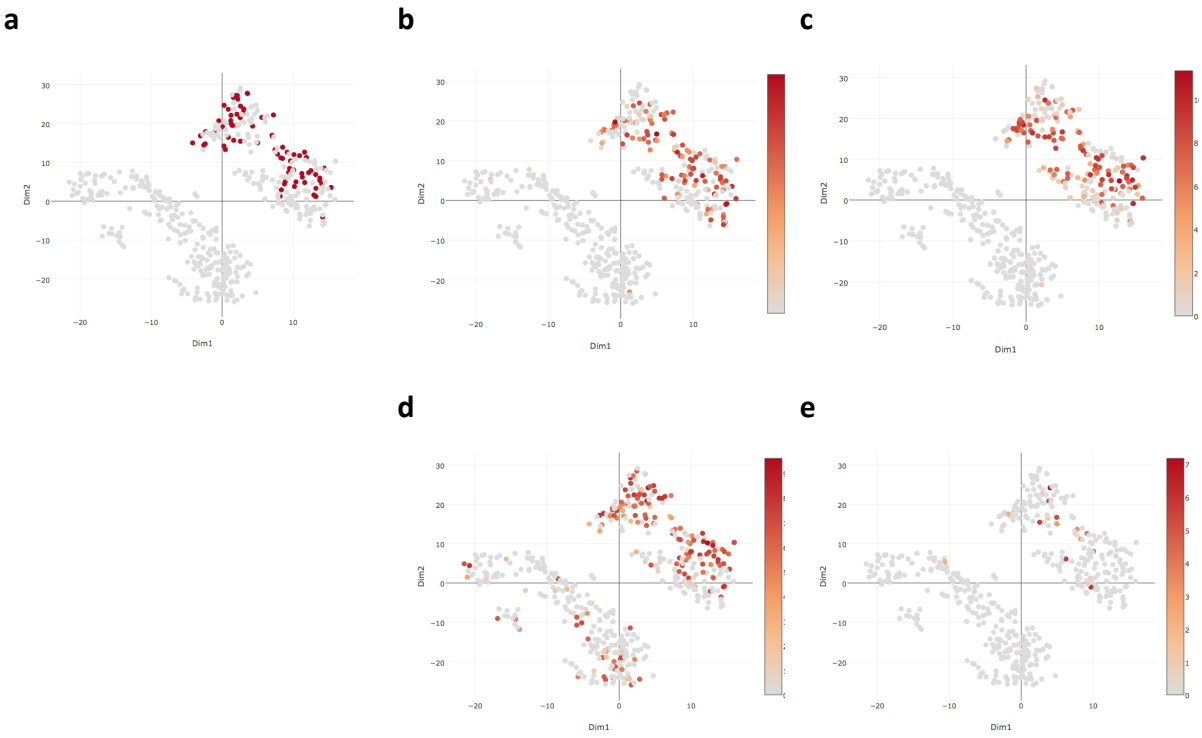


Figure 4.7 Gene expression signatures correlate with DC subtypes. (a) EV-transduced Hoxb8-FL cells after 5 days of differentiation in Flt3L. (b) (c) (d) and (e) show gene expression of cd300c, Siglec-H, Slc46a3 and Tspan33 respectively at Day 5 of differentiation.

CD300c and the canonical pDC gene Siglec-H are both found in module F150 in the ImmGen Consortium database and show convincing gene expression profiles which correlate with EV-transduced Hoxb8-FL cells differentiated in Flt3L. The genes Slc46a3 and Tspan33 are less well-established markers for DC expression but are derived from the ImmGen database F156 module and show some correlation with Day 5 differentiated cells. By contrast, none of the genes in the F152 module are highly expressed in any of the Hoxb8-FL transduced cell lines, confirming our preliminary findings from XCR1 antigen-negativity that CD8α^+ type 1 cDCs are not generated by the differentiation of Hoxb8-FL cells.

For over 10 years, *IRF8* and *IRF4* expression have been known to distinguish between DC subtypes (Tamura *et al*, 2005). *IRF8* expression is elevated in both pDCs and CD8α^+ type 1 cDCs, whereas *IRF4* expression has been thought to be elevated in type 2 cDCs. Interestingly, our gene expression data suggests that Hoxb8-parental cells differentiated in Flt3L express *IRF8* and lack expression of *IRF4* (Figure 4.8):

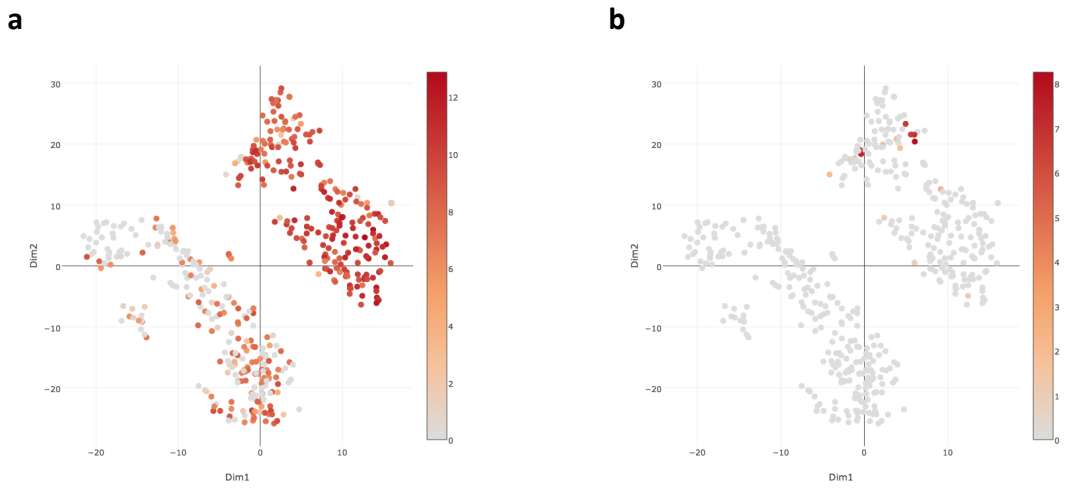


Figure 4.8 ASAP shows (a) IRF8 and (b) IRF4 expression of transduced Hoxb8-parental cells after 5 days of differentiation in Flt3L

This seems to contradict our previous findings on flow cytometry that Hoxb8-parental cells in Flt3L differentiate into CD11b$^+$ SIRPα^+ type 2 cDCs. However, it is important to recognise that our understanding of DC lineages and gene expression profiles is rapidly evolving. For instance, recent work has demonstrated that type 2 cDCs in humans are heterogeneous and include a subpopulation of CD5^{low} DCs which demonstrate decreased expression of *IRF4* (Yin *et al*, 2017). The co-expression of CD11c, CD11b and SIRPα in both our flow cytometry and our scRNA-seq studies is strongly indicative that transduced Hoxb8 cells exhibit a type 2 cDC phenotype at Day 5 of differentiation, whereas the expression of *IRF8* in the absence of XCR1 suggests the presence of pDCs.

4.4.4 LMPP-like cells transduced with *CEBPA* N321D lose the ability of *CEBPA* WT-transduced cells to overcome Hox-mediated differentiation arrest

Gain-of-function studies with overexpression of Hox genes have indicated that the Hox family of transcription factors expand haematopoietic stem and progenitor cell (HSPC) populations and arrest lineage differentiation programs (Pineault *et al*, 2002, Fischbach *et al*, 2005, Spencer *et al*, 2015, Brunetti *et al*, 2018). Hoxb8-FL parental cells are conditionally

immortalised in Flt3L by estradiol-mediated translocation of Hoxb8 from the cytoplasm to the nucleus where it can bind to and transcribe DNA. Previous work on Hoxb8-FL cells shows that in the presence of Flt3L and estradiol these cells are negative for Gr1, CD11b, B220 and MHC class II (Redecke *et al*, 2013).

Differential expression was performed on cells that passed QC using ASAP's DESeq2 package between *CEBPA* WT and EV-transduced cells at Day 2 post-transduction, with an FDR (adjusted p-value) ≤ 0.05 and a fold change ≥2. It is well-established that *CEBPA* expression is required for granulocyte differentiation at the CMP to GMP junction (Zhang *et al*, 2004). Interestingly, the integration of exogenous human *CEBPA* WT into Hoxb8-FL cells causes a dramatically altered transcriptional profile at an early LMPP-like stage with 636 upregulated and 429 downregulated genes (Appendix, Tables 1 and 3 respectively).

Enrichr (http://amp.pharm.mssm.edu/Enrichr/) is an online interactive gene list enrichment analysis tool designed by the Ma'ayan group (Chen *et al*, 2013, Kuleshov *et al*, 2016). Genes differentially upregulated in *CEBPA* WT-transduced cells express both neutrophilic and antigen presentation ontogenies when analysed on Enrichr, for instance GO biological processes are characterised by neutrophil degranulation and neutrophil activation, but also by TLR4 signalling (Appendix, Table 2).

The 636 upregulated genes include genes which are known to play significant roles in granulocyte-monocyte characterisation and function, including *CD33*, *GCA*, *CHIT1*, *LY6C*, *SIRPA*, and *ITGAM* (Ripperger *et al*, 2015). Enrichr analysis of the 429 downregulated genes suggests enrichment in cell ontologies on Mouse gene atlas which include HSCs, MEPs and CMPs (Appendix, Table 4). This confirms that the EV-transduced cells remain in a primitive state in the presence of estradiol, whereas the *CEBPA* WT-transduced cells overcome Hoxb8-mediated differentiation block and progress towards differentiation in the granulocyte lineage. In addition, these results signify that it is difficult to compare the trajectories between Days 2 and 7 of EV or *CEBPA* N321D-transduced cells with the trajectory between Days 2 and 7 of *CEBPA* WT-transduced cells, because the *CEBPA* WT-transduced cells start with very different transcriptional and phenotypic profiles at Day 2.

By contrast, the transcriptional signatures of EV and *CEBPA* N321D-transduced cells are remarkably similar to each other at Day 2 post-transduction before the removal of estradiol, as demonstrated by the low number of genes (15 upregulated and 5 downregulated) differentially expressed by *CEBPA* N321D-transduced single cells versus EV-transduced cells at this time point (see Appendix, Table 5). Unsurprisingly, differential expression of *CEBPA* WT versus *CEBPA* N321D-transduced cells at Day 2 gives 745 genes which suggest enrichment for GO biological processes including neutrophil degranulation and 292 downregulated genes which Enrichr analysis correlates with HSPC ontologies. This confirms and further characterises the transcriptional signatures visualised in t-SNE, that exogenous *CEBPA* WT is able to overcome Hoxb8-mediated maturation arrest by downregulating expression of HSPC-associated genes and upregulating genes associated with neutrophil cell ontology, and furthermore that N321D-mutated *CEBPA* cannot replicate these WT transcriptional effects in the presence of Hoxb8 nuclear translocation.

4.4.5 Differential expression identifies genes up and downregulated in EV, *CEBPA* WT and *CEBPA* N321D-transduced cells during differentiation in Flt3L

Differential expression was then performed on the cells which passed QC using ASAP's DESeq2 package on each cell line between Day 7 and Day 2 post-transduction, with an FDR (adjusted p-value) ≤ 0.05 and a fold change ≥ 2. The results of differential expression can be found in Table 4.2:

Table 4.2 Differential expression for each cell line during 5 days in differentiation media (Flt3L without estradiol)

	Upregulated genes	Downregulated genes
EV-transduced cells	1400	1340
CEBPA WT-transduced cells	1167	738
CEBPA N321D-transduced cells	594	418

The *CEBPA* N321D-transduced cells clearly have the lowest number of upregulated and downregulated genes during 5 days of differentiation. Interestingly, *CEBPA* WT-transduced cells have a lower number of genes differentially expressed than EV-transduced cells, but it should be remembered that flow cytometry and differential expression data indicated evidence of differentiation before the removal of estradiol, and this may account for the more subdued transcriptional trajectory seen after the removal of estradiol.

The altered expression signature in *CEBPA* WT cells at Day 2 post-transduction is also important because these cells start with a different transcriptional status to the EV or *CEBPA* N321D-transduced cells, and therefore genes which are differentially expressed in *CEBPA* WT-transduced cells are not directly comparable to genes derived from differential expression of the other subsets.

4.4.6 Gene enrichment suggests pDC gene ontology for *CEBPA* N321D-transduced cells differentiated in Flt3L

ASAP has several gene enrichment tools based on its own unique enrichment method which is based on a Fisher Exact Test. By default, it enriches the top 100 genes ranked by fold-change. The following gene enrichment data is acquired by enriching the top 100 upregulated genes identified by differential expression between datasets from *CEBPA* N321D-transduced cells at Days 7 and 2 (Table 4.3):

Table 4.3 Gene enrichment of top 100 genes ranked by fold-change provides gene atlas and gene ontology for CEBPA N321D-transduced cells

Gene Atlas		
Annotation	OR	Adjusted p-value
dendritic_plasmacytoid_B220+	11.59893455	0.000106858
macrophage_peri_LPS_thio_0hrs	5.723928707	0.006772816
RAW_264_7	7.00752457	0.009999595
macrophage_bone_marrow_0hr	7.106759907	0.021267311
macrophage_peri_LPS_thio_1hrs	3.756913022	0.037620484

GO Biological Processes			
Annotation	Description	OR	Adjusted p-value
GO:0002250	adaptive immune response	9.647636039	0.000107661
GO:0019221	cytokine-mediated signalling pathway	9.524060414	0.000595797
GO:0002449	lymphocyte mediated immunity	10.82278481	0.000595797
GO:0006935	chemotaxis	7.013919095	0.000595797
GO:0042330	taxis	6.984405458	0.000595797
GO:0071345	cellular response to cytokine stimulus	7.043805165	0.001256338
GO:0002443	leukocyte mediated immunity	8.290459604	0.002432459
GO:0048002	antigen processing and presentation of peptide antigen	22.77316161	0.002960939
GO:0009615	response to virus	7.37755102	0.013946116
GO:0060326	cell chemotaxis	8.471471471	0.016538952
GO:0002495	antigen processing and presentation of peptide antigen via MHC class II	48.45882353	0.021975521

GO:0050778	positive regulation of immune response	5.779032018	0.021975521
GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	45.60622837	0.022476632
GO:0030595	leukocyte chemotaxis	9.50174216	0.022476632
GO:1903708	positive regulation of haemopoiesis	9.225010574	0.024244182
GO:0019882	antigen processing and presentation	12.31674361	0.024244182
GO:0051607	defense response to virus	7.220702754	0.025565976
GO:0050900	leukocyte migration	7.137351547	0.025883188
GO:0002573	myeloid leukocyte differentiation	8.6218853	0.027668663
GO:0002478	antigen processing and presentation of exogenous peptide antigen	32.29411765	0.037621686

GO Molecular Functions			
Annotation	Description	OR	Adjusted p-value
GO:0004896	cytokine receptor activity	27.61865086	2.50E-06
GO:0042287	MHC protein binding	73.19444444	1.68E-05
GO:0019955	cytokine binding	20.93680556	4.78E-05

Interestingly in the context of *CEBPA*'s role in myeloid differentiation, myeloid leukocyte differentiation is identified as one of the biological processes in *CEBPA* N321D-transduced cells after 5 days of differentiation in Flt3L. Nevertheless, ASAP's gene enrichment tools all coincide to suggest a cell ontology which is characterised by mechanisms of the adaptive immune response such as cytokine binding and antigen presentation. Gene Atlas annotation suggests that these cells correlate with the gene expression profiles of $B220^{+}$ pDCs. Therefore it appears that the *CEBPA* N321D mutation switches on a pDC transcriptional program in Hoxb8-FL cells as early as Day 5 of differentiation in Flt3L, although phenotypic cell surface markers cannot be detected by flow cytometry until much later.

Genes of particular interest are the genes up and downregulated during differentiation (between Days 7 and 2 post-transduction) of *CEBPA* N321D-transduced cells, which are not identified during differentiation of EV-transduced cells. Bearing in mind that the *CEBPA* WT trajectory is not directly comparable to the trajectories of EV or *CEBPA* N321D-transduced cells during differentiation in Flt3L, for reasons aforementioned, these genes of interest should not take into consideration genes differentially expressed in *CEBPA* WT-transduced cells between Days 7 and 2. Differentially expressed genes can be intersected by using the online tool http://bioinformatics.psb.ugent.be/webtools/Venn (Figure 4.9):

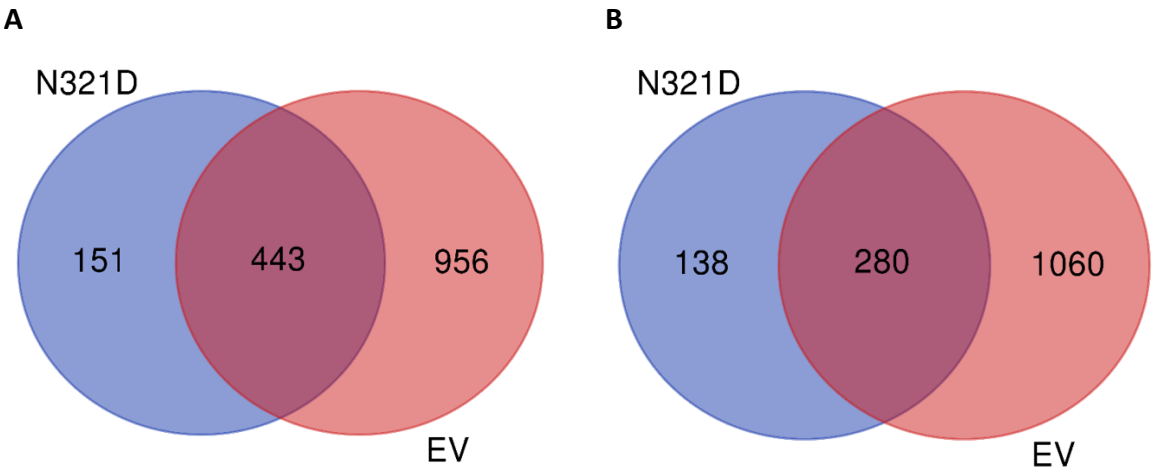


Figure 4.9 Number of genes differentially expressed for cell lines differentiated for 5 days in Flt3L, and not taking into consideration genes differentially expressed in *CEBPA* WT-transduced cells. (A) Upregulated genes, (B) Downregulated genes. (Appendix, Tables 6 and 7 respectively).

ASAP's gene enrichment tools allow only global gene set analysis based on differential expression between defined clusters of cells. Furthermore, it does not provide the identities of specific genes. Therefore, web-based Enrichr gene set enrichment analysis was performed on the 151 genes identified as differentially upregulated exclusively in the *CEBPA* N321D-transduced cells. Enrichr calculates a Combined Score by taking the log of the p-value from the Fisher Exact Test and multiplying it by the z-score of the deviation from the expected rank:

Table 4.4 Enrichr calculates Combined Score and identifies upregulated genes which overlap with top ten entries in Mouse Gene Atlas after differential expression analysis of *CEBPA* N321D-transduced cells at Day 7 v Day 2

Term	Combined Score	Genes
NK_cells	10.16681656	C330007P06RIK;NOTCH1;NUP210;OSBPL3;GAB3;LNPEP;TM6SF1;SSH2;ZEB2;1700025G04RIK;KLRB1C;BCL2;FAM65B;FYN
dendritic_plasmacytoid_B220+	6.233181251	SLC44A1;TRAPPC5;MGAT1;FNDC3A;IFFO1;RUNX2;RAB33B
mast_cells_IgE+antigen_1hr	3.279761749	D8ERTD82E;GFOD1;CD93;SLC44A1
Baf3	2.537389426	FBXW4;EGFL7;BCL2;2310001H17RIK;CLEC2I;CBFA2T3;OCEL1
macrophage_bone_marrow_0hr	2.523596016	DGKD;CD93;TM6SF1;DUSP7
dendritic_cells_myeloid_CD8a-	2.451664086	TIFAB;TRAPPC5;BMF;CBFA2T3
mammary_gland_non-lactating	2.133937578	MYLPF;TPM2;SGCG
follicular_B-cells	1.916955745	MEF2C;SNX29;DGKD;FAM65B;FAM76B;RPRD2;AVL9;TRPC4AP;FLI1;D930015E06RIK;CMAH
granulo_mono_progenitor	1.370213386	TIFAB;KCNK12
mast_cells	1.284288746	IVNS1ABP;SCIN;BCL2;GAB3;SERINC3;CBFA2T3;FLI1

This analysis confirms a pDC cell ontology in *CEBPA* N321D-transduced cells differentiated for 5 days in Flt3L, and identifies 5 upregulated genes which are associated with the pDC phenotype: *SLC44A1*,*TRAPPC5*, *MGAT1*, *FNDC3A*, *IFFO1*, *RUNX2*, and *RAB33B*. Interestingly, it also suggests that differentiation of *CEBPA* N321D-transduced cells upregulates genes (*TIFAB* and *KCNK12*) associated with GMP cell ontology.

4.4.7 Gene enrichment suggests that Empty Vector and *CEBPA* WT-transduced cells have gene ontologies which are distinct from *CEBPA* N321D-transduced cells differentiated in Flt3L

Enrichr analysis of the 956 genes exclusively upregulated in EV-transduced cells during differentiation in Flt3L suggests a distinct gene ontology:

Table 4.5 Enrichr calculates Combined Score to identify upregulated genes which overlap with top ten entries in Mouse Gene Atlas after differential expression analysis of EV-transduced cells at Day 7 v Day 2

Term	Combined Score	Genes
macrophage_peri_LPS_thio_7hrs	113.42	134
dendritic_cells_myeloid_CD8a-	15.67	32
macrophage_peri_LPS_thio_0hrs	14.73	56
follicular_B-cells	9.52	58
macrophage_bone_marrow_6hr_LPS	9.25	80
macrophage_peri_LPS_thio_1hrs	7.85	59
macrophage_bone_marrow_24h_LPS	7.63	58
macrophage_bone_marrow_0hr	6.97	31
macrophage_bone_marrow_2hr_LPS	4.44	45
dendritic_cells_lymphoid_CD8a+	4.27	22

Furthermore, a direct comparison between cell ontologies on Mouse Gene Atlas of genes downregulated exclusively during differentiation of EV and *CEBPA* N321D-transduced cell lines in Flt3L demonstrates distinct differential expression signatures (Table 4.6).

Table 4.6 Enrichr identifies Mouse Gene Atlas cell ontologies with Combined Scores for downregulated genes after differential expression analysis of EV and *CEBPA* N321D-transduced cells at Day 7 v Day 2

Empty Vector-transduced cells		*CEBPA* N321D-transduced cells	
Term	Combined Score	Term	Combined Score
embryonic_stem_line_V26_2_p16	13.79	mast_cells_IgE+antigen_1hr	5.82
embryonic_stem_line_Bruce4_p13	11.62	thymocyte_SP_CD4+	3.04
mega_erythrocyte_progenitor	9.65	bone_marrow	2.89
Baf3	6.01	bladder	2.46
stem_cells__HSC	4.02	macrophage_peri_LPS_thio_0hrs	2.41
nih_3T3	2.99	embryonic_stem_line_Bruce4_p13	2.36
common_myeloid_progenitor	1.73	mast_cells	1.71
granulo_mono_progenitor	1.05	mega_erythrocyte_progenitor	1.62
mIMCD-3	0.43	macrophage_peri_LPS_thio_1hrs	1.58
adipose_brown	0.32	mast_cells_IgE+antigen_6hr	1.45

Genes up and downregulated during differentiation of *CEBPA* WT-transduced cells in Flt3L should be considered separately from gene sets identified in EV and *CEBPA* N321D-transduced cells, as their differentiation trajectories start from a different transcriptional state at Day 2 post-transduction. Gene set enrichment analysis by Enrichr of 1167 upregulated and 738 downregulated genes identified by differential expression analysis of *CEBPA* WT-transduced cells at Days 7 and 2 in Flt3L demonstrate the following cell ontologies in Mouse Gene Atlas:

Table 4.7 Enrichr identifies Mouse Gene Atlas cell ontologies with Combined Scores for up and downregulated genes identified by differential expression analysis *CEBPA* WT-transduced cells at Day 7 v Day 2

Upregulated genes		Downregulated genes	
Term	Combined Score	Term	Combined Score
macrophage_peri_LPS_thio_7hrs	113.65	adipose_brown	8.38
follicular_B-cells	23.03	bone	8.01
macrophage_bone_marrow_6hr_LPS	17.12	stem_cells__HSC	6.2
dendritic_cells_myeloid_CD8a-	16.56	bone_marrow	4.41
macrophage_bone_marrow_0hr	14.85	macrophage_bone_marrow_24h_LPS	3.36
macrophage_peri_LPS_thio_0hrs	12.28	granulo_mono_progenitor	2.64
macrophage_peri_LPS_thio_1hrs	9.46	granulocytes_mac1+gr1+	1.87
dendritic_plasmacytoid_B220+	9.2	3T3-L1	1.76
macrophage_bone_marrow_2hr_LPS	5.38	Baf3	1.24
RAW_264_7	4.26	uterus	1.22

In summary, gene set enrichment analysis for upregulated genes differentially expressed during differentiation of EV-transduced cells in Flt3L gives cell ontologies which differ markedly from *CEBPA* N321D-transduced cells (Table 4.4), with prominence of cDC type 2 cell ontology and the exclusion of $B220^{+}$ pDC cells from the top ten statistically significant cell ontologies (Table 4.5). Gene enrichment analysis from differential expression of *CEBPA* WT-transduced cells needs to be interpreted with more caution because their differentiation trajectory starts from a distinct transcriptional state, but upregulated genes demonstrate a similar cell ontology to EV-transduced cells, and genes associated with myeloid CD8α- DCs have higher statistical significance than genes associated with $B220^{+}$ pDCs (Table 4.7). Perhaps more significantly from the perspective of the leukaemic phenotype demonstrated by the *CEBPA* N321D mutation *in vivo*, the *CEBPA* N321D-transduced cells fail to downregulate gene sets associated with HSPC cell ontology which are downregulated by both EV (Table 4.6) and *CEBPA* WT (Table 4.7) -transduced cells.

4.4.8 Intersection of differentially expressed genes in N321D and EV subsets suggests transcriptional reprogramming activated by mutated *CEBPA*

Phenotypically, EV-transduced cells proceed rapidly to terminal differentiation whereas *CEBPA* N321D-transduced cells do not acquire features of terminal differentiation and can be cultured for long periods of time in the presence of only Flt3L. It would therefore be useful to identify genes which are upregulated exclusively in EV-transduced cells after 5 days of differentiation in Flt3L (956 genes) and intersect those genes with the 633 genes which are downregulated on differential expression between *CEBPA* N321D and EV-transduced cells at Day 7. This would provide us with a list of genes that characterise the differentiation of EV cells and which fail to be upregulated in *CEBPA* N321D-transduced cells.

Remembering that EV-transduced cells represent the normal pattern of differentiation and lineage restriction by LMPP-like Hoxb8-FL cells, it is also useful to identify the 1060 genes downregulated exclusively in EV cells during differentiation in Flt3L and intersect them with the 224 genes which are upregulated on differential expression between *CEBPA* N321D and EV-transduced cells at Day 5.

By intersecting these data subsets, we are able to construct a list of 385 EV differentiation genes which are switched off by *CEBPA* N321D-transduction (Appendix, Table 8), and

alternatively a list of 133 genes switched off in differentiated EV-transduced cells which are switched on by *CEBPA* N321D-transduction (Appendix, Table 8). These two gene sets can then be analysed by Enrichr:

Table 4.8 Enrichr identifies Mouse Gene Atlas cell ontologies with Combined Scores for genes "switched on" and "switched off" after intersection of gene subsets extracted from differential expression of *CEBPA* N321D and EV-transduced cells

Switched on genes		Switched off genes	
Term	Combined Score	Term	Combined Score
placenta	3.04	macrophage_peri_LPS_thio_7hrs	77.87
common_myeloid_progenitor	2.47	macrophage_bone_marrow_6hr_ LPS	24.95
mega_erythrocyte_progenitor	2.18	macrophage_peri_LPS_thio_1hrs	15.05
Baf3	2.1	macrophage_bone_marrow_0hr	11.91
stem_cells__HSC	1.58	dendritic_cells_myeloid_CD8a-	7.79
thymocyte_SP_CD4+	1.5	follicular_B-cells	5.31
mast_cells	1.13	macrophage_bone_marrow_24h _LPS	4.64
nih_3T3	0.92	microglia	4.5
osteoblast_day21	0.72	osteoclasts	3.67
osteoblast_day14	0.6	macrophage_peri_LPS_thio_0hrs	3.48

Enrichr cell ontology of *CEBPA* N321D-mediated "switched off" genes (Table 4.8) suggests that mutant *CEBPA* dedifferentiates macrophages and myeloid dendritic cell or cDC ontology, but not pDC ontology. Interestingly, some of these genes encode known partner TFs of wildtype C/EBPα such as *C-FOS* and *C-JUN* (Hong *et al*, 2011). By contrast, "switched on" genes include *CDK6* and give cell ontology of HSCs, MEPs and CMPs on Enrichr.

Essentially, *CEBPA* N321D-transduced cells differentiated at Day 5 in the presence of Flt3L seem to "switch on" or fail to switch off a HSPC transcriptional program. Ideally this intersectional analysis should include *CEBPA* WT-transduced cells to demonstrate that these two transcriptional programs are derived specifically from the N321D mutation rather than from overexpression of *CEBPA* WT, but this is difficult due to the altered transcriptional state of *CEBPA* WT-transduced cells at Day 2 post-transduction. However, phenotypic features of *CEBPA* WT-transduced cells, including cell growth and cell surface markers, strongly suggest that the HSPC transcriptional program is specific to the N321D mutation.

4.4.9 A small subset of *CEBPA* N321D-transduced cells demonstrate a divergent transcriptional signature on differential expression against EV-transduced cells

t-SNE visualisation of single-cell expression profiles reveals that the majority of the *CEBPA* N321D-transduced cells cluster proximal to the Day 5 differentiated EV and *CEBPA* WT-transduced cells after five days of differentiation in Flt3L. The most significant finding on the level of the single cell is that a rare subpopulation of 5 *CEBPA* N321D-transduced cells diverge from the Day 5 differentiated populations entirely and cluster with the Day 0 non-differentiated populations:

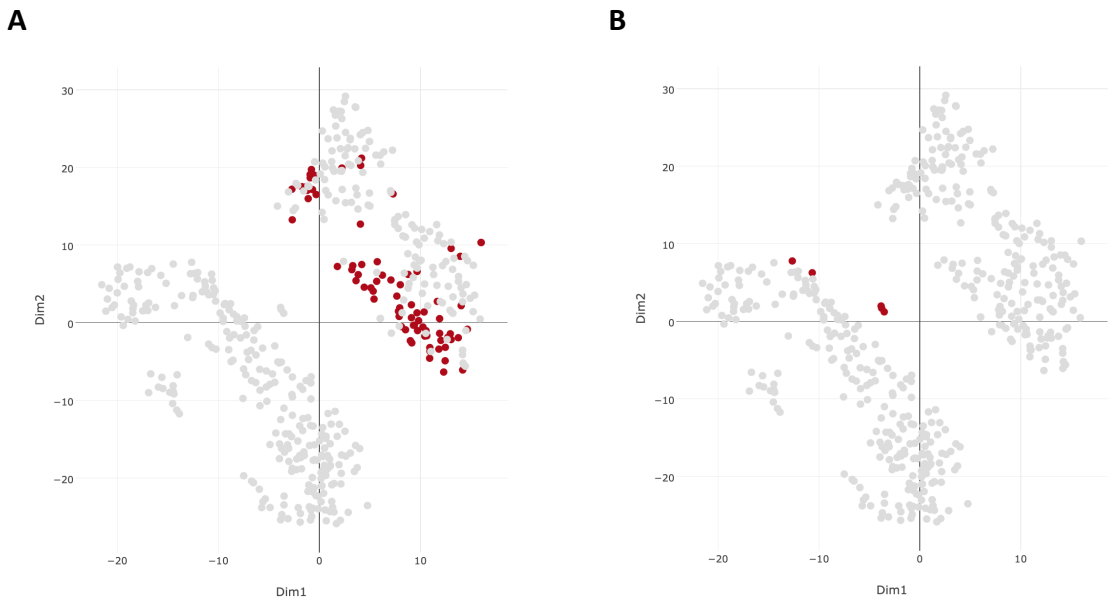


Figure 4.10 ASAP shows two cohorts of *CEBPA* N321D-transduced cells after differentiation for 5 days in Flt3L. (A) 78 single cells which are distributed proximally but close to EV and WT-transduced cells differentiated in Flt3L, and (B) 5 single cells which diverge from Day 5 populations and cluster close to the Day 0 non-differentiated populations

Differential expression with DESeq2 confirms that the two cohorts of *CEBPA* N321D-transduced cells differentiated at Day 5 express divergent transcriptional signatures. Genes upregulated and downregulated in the 78 and 5 cells were identified by applying DESeq2 separately to cells from each cohort against the 82 *CEBPA* N321D-transduced cells at Day 0. The differentially expressed genes were then intersected with genes differentially expressed between EV-transduced cells at Days 5 and 0 (Figure 4.11):

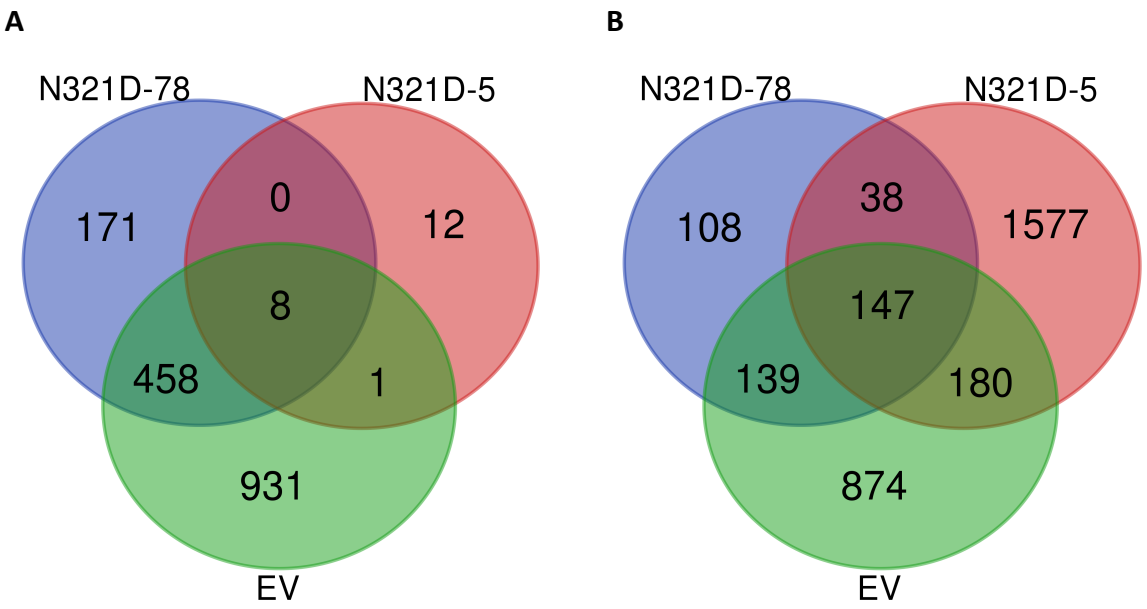


Figure 4.11 Number of genes differentially expressed between Day 5 versus Day 0 for two cohorts of 78 and 5 *CEBPA* N321D-transduced and all EV-transduced single cells differenti ated inFlt3L, and not taking into consideration genes differentially expressed in *CEBPA* WT-transduced cells. (A) Upregulated genes, (B) Downregulated genes. See Appendix Tables 9 and 10 for full lists of genes.

Enrichr analysis was then performed for genes which were upregulated (171 and 12) and downregulated (108 and 1577) respectively in the 78 (Table 4.9) and 5 (Table 4.10) single *CEBPA* N321D-transduced cells:

Table 4.9 Enrichr identifies Mouse Gene Atlas cell ontologies with Combined Scores for up and downregulated genes identified by differential expression analysis of 78 *CEBPA* N321D-transduced cells at Day 7 v all 82 *CEBPA* N321D-transduced cells at Day 2, excluding genes differentially expressed in EV-transduced cells between the same time points

Upregulated genes (171)		Downregulated genes (108)	
Term	Combined Score	Term	Combined Score
follicular_B-cells	6.90	mast_cells_IgE+antigen_1hr	5.10
NK_cells	6.34	embryonic_stem_line_Bruce4_p13	3.17
dendritic_plasmacytoid_B220+	3.54	stem_cells__HSC	2.11
mast_cells_IgE+antigen_1hr	2.62	thymocyte_SP_CD4+	1.98
osteoblast_day5	2.13	macrophage_peri_LPS_thio_7hrs	1.89
macrophage_bone_marrow_0hr	2.05	embryonic_stem_line_V26_2_p16	1.78
dendritic_cells_myeloid_CD8a-	1.97	B-cells_GL7_positive_KLH	1.77
Baf3	1.78	lens	1.74
mammary_gland_non-lactating	1.70	mega_erythrocyte_progenitor	1.65
mast_cells	1.68	bladder	1.52

Table 4.10 Enrichr identifies Mouse Gene Atlas cell ontologies with Combined Scores for up and downregulated genes identified by differential expression analysis of 5 *CEBPA* N321D-transduced cells at Day 7 v all 82 *CEBPA* N321D-transduced cells at Day 2, excluding genes differentially expressed in EV-transduced cells between the same time points

Upregulated genes (12)		Downregulated genes (1577)	
Term	Combined Score	Term	Combined Score
microglia	3.27	bone_marrow	1.43
macrophage_peri_LPS_thio_0hrs	0.00	follicular_B-cells	1.41
macrophage_bone_marrow_24h _LPS	0.00	macrophage_peri_LPS_thio_0hrs	0.84
		adipose_brown	0.81
		B-cells_GL7_positive_Alum	0.69
		B-cells_GL7_positive_KLH	0.60
		mast_cells	0.50
		B-cells_GL7negative_Alum	0.50
		mast_cells_IgE	0.45
		spleen	0.38

The 78 single *CEBPA* N321D-transduced cells which cluster within the proximal aspect of the Day 5 differentiated cells upregulate genes which enrich for mononuclear phagocyte cell ontologies including pDCs, cDCs, macrophages and mast cells, and downregulate genes which enrich for HSC ontologies. By contrast, the 5 single cells which diverge from the Day 5 differentiated cell populations do not display this transcriptional signature. In fact, the 5 single cells upregulate only 12 genes compared to the Day 0 cells, after excluding genes upregulated by EV during differentiation (Figure 4.11). Interestingly, although they downregulate 1577 genes, the low Combined Scores of enriched cell ontologies associated with those 1577 genes (Table 4.10) suggests that the enriched gene sets are relatively non-specific.

Experimentally, the small cluster of 5 single cells all originate from one single plate which raises questions of bias from a plate effect, however the majority of *CEBPA* N321D-transduced cells from the same plate cluster with cells from other plates. Bearing in mind that the Day 0 Hoxb8-FL cells, in the presence of Flt3L and estradiol-mediated Hoxb8 DNA-binding, are conditionally immortalised and capable of exponential growth, one can hypothesise that it is in fact the rare smaller population of 5 *CEBPA* N321D-transduced single cells which are immortalised when estradiol is withdrawn and enjoy a selective survival advantage. This would account for the gradual evolution of the cell surface marker profile of the mutant cells

from an initial obvious reduction in CD11b and SIRPα expression to a $CD11c^+$ $B220^+$ cell line characterised by exponential growth.

4.4.10 Validation of ASAP gene expression data using geometric size factor normalisation and ggplot2

To validate gene expression data from ASAP, it was decided to use the function estimateSizeFactorsForMatrix (R version 3.5.2, DESeq2 version 1.22.2) on genomic data for cells which passed QC. GFP expression data was specifically removed before calculating size factor as it would skew normalisation. Size factors were then applied to raw counts to normalise them, before log transforming the results and filtering out zero counts. The R-package ggplot2 version 3.1.0 was then employed to construct violin plots and compare gene expression data for EV, *CEBPA* WT and *CEBPA* N321D-transduced conditions at Days 2 and 7:

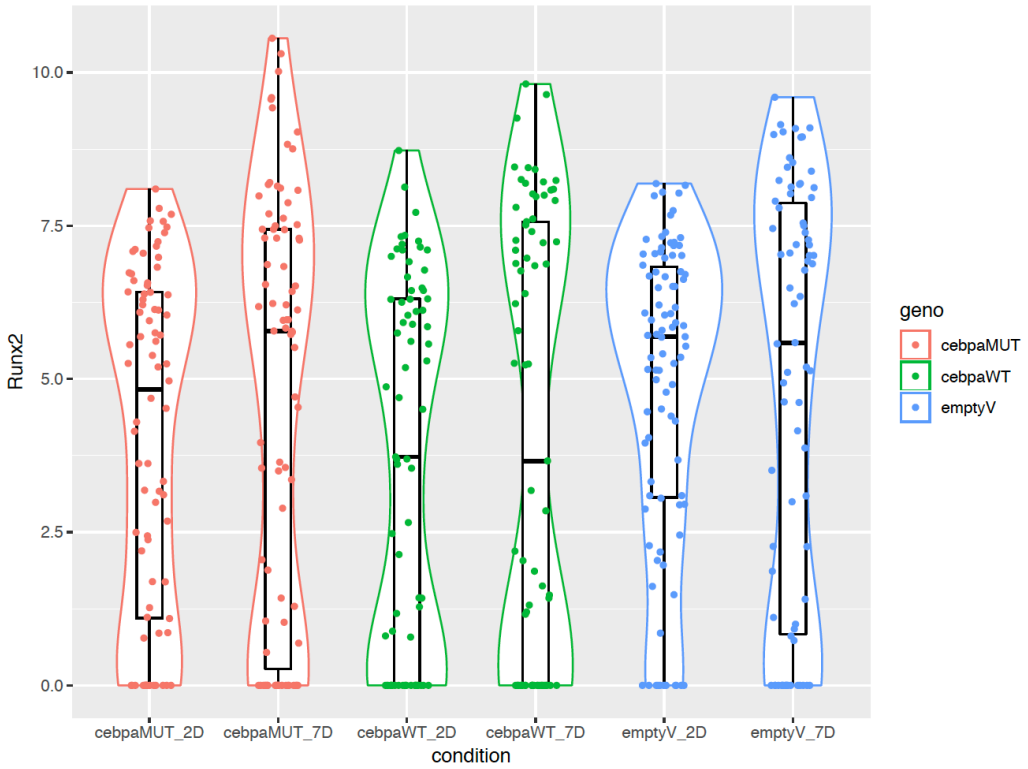


Figure 4.12 Violin plot showing normalised counts for *RUNX2* in all cells which passed QC

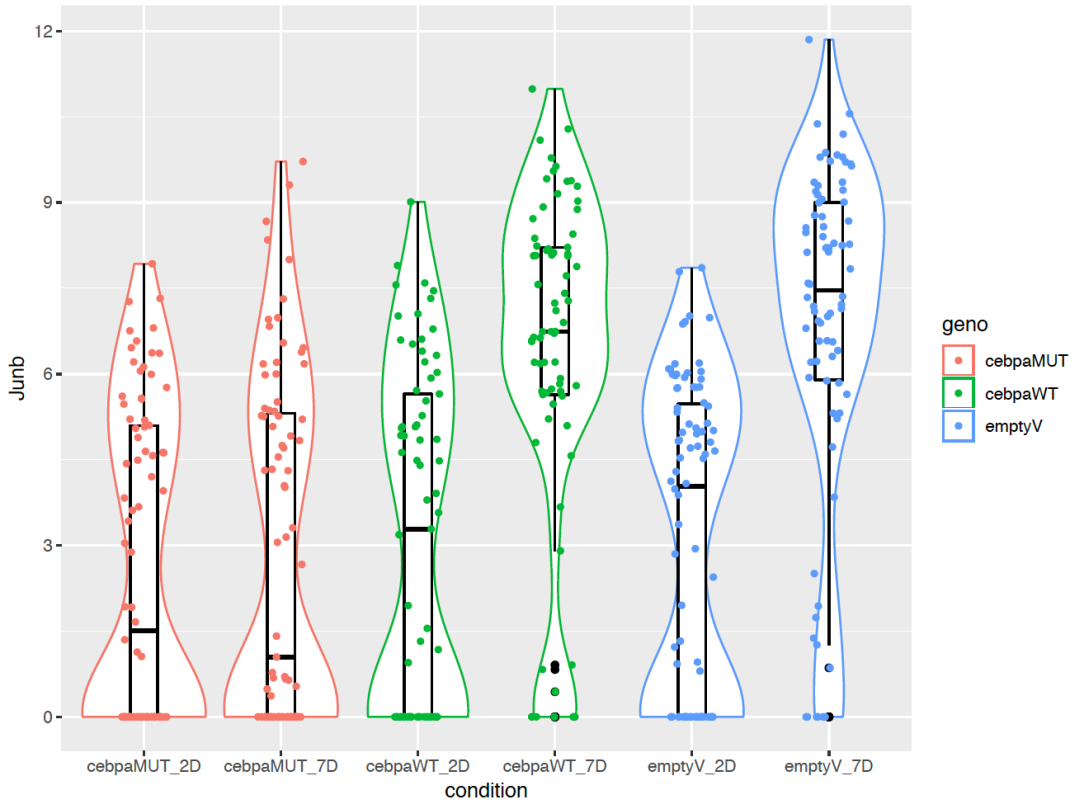


Figure 4.13 Violin plot showing normalised counts for *JUNB* in all cells which passed QC

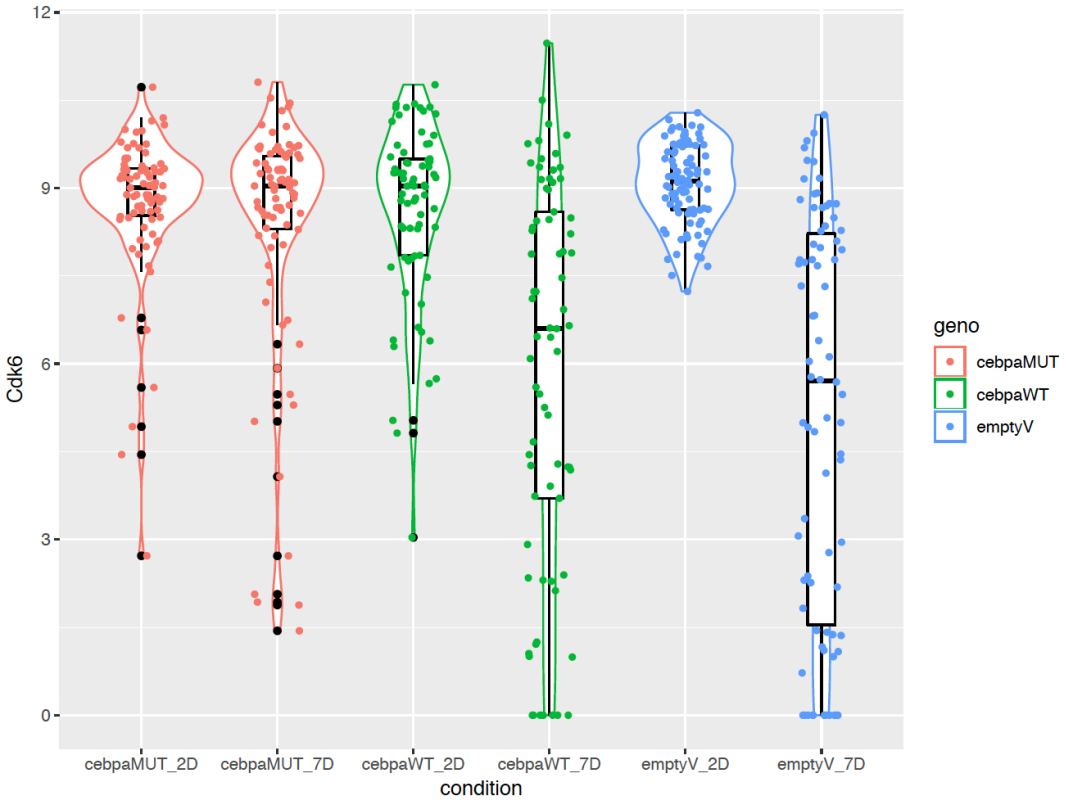


Figure 4.14 Violin plot showing normalised counts for *CDK6* in all cells which passed QC

These violin plots provide convincing visual validation that *RUNX2* is upregulated exclusively in *CEBPA* N321D-transduced cells, that *CEBPA* N321D "switches off" genes associated with macrophage programs such as *JUNB*, and that it "switches on" genes including *CDK6* which give cell ontology of early haematopoietic precursors.

4.5 Discussion

My cellular model is designed to evaluate early pre-leukaemic changes induced by the *CEBPA* N321D mutation, by employing Hoxb8-FL cells which are phenotypically similar to LMPP cells and therefore present an opportunity to characterise *CEBPA*-mutated cells before the CMP to GMP junction where *CEBPA* is known to be highly expressed.

The original rationale for examining *CEBPA* WT-transduced cells was to provide a comparator to exclude effects from over-expression of *CEBPA*, such that genes dysregulated exclusively by *CEBPA* mutation can be identified. This comparison was complicated by the fact that *CEBPA* WT-transduced cells transition into a distinct transcriptional state in Hoxb8-FL conditions. The rationale for performing the single-cell experiment was that it enables the identification of rare cell populations and cell-to-cell variations, which are important when studying early haematopoiesis or pre-leukaemic transcriptional signatures. To our knowledge, the gene expression signature of *CEBPA*-mutated cells has never previously been evaluated at the level of the single cell.

4.5.1 Role of C/EBPα in myeloid differentiation and DC biology, and the single-cell expression profile of *CEBPA* WT-transduced Hoxb8-FL cells in conditions which favour immortalisation

C/EBPα plays a key role in myeloid differentiation, and is able to cause growth arrest and myeloid differentiation when lentivirally inserted into $CD34^+$ AML cells (Schepers *et al*, 2007). It is expressed at a low level in HSCs, incrementing at the stage of pre-GPs or CMPs, and then increasing markedly in GMPs (Zhang *et al*, 2004).

The role of C/EBPα in HSC biology is complex, with the Tenen group reporting that deletion of C/EBPα increases HSC proliferation and promotes HSC expansion through an N-Myc mechanism (Ye *et al*, 2013). By contrast, the Porse group found that loss of C/EBPα resulted

in HSC exhaustion and a marked reduction of haematopoietic reconstitution after competitive transplantation into lethally irradiated recipient mice (Hasemann *et al*, 2014). The apparent contradiction in these findings may relate to the timing of C/EBPα excision in these two experimental models, and it is possible to integrate these apparently divergent results by hypothesising early HSC expansion followed by HSC exhaustion and loss of self-renewal.

Regarding lineage restriction, initial studies suggested that C/EBPα was required for the transition from CMPs to GMPs, based on the observation that Sca-1^{-} progenitor cells accumulated in adult *CEBPA* $^{-/-}$ mice (Zhang *et al*, 2004, Hasemann *et al*, 2014). However, gene expression data has recently mapped *CEBPA* $^{-/-}$ pre-GMP cells more closely to LSK cells, suggesting that the maturation block caused by *CEBPA* deletion lies just downstream of LSK cells when Sca-1 expression is downregulated (Pundhir *et al*, 2018).

Both HSC self-renewal (Hasemann *et al*, 2014) and myeloid differentiation are mediated by C/EBPα acting in partnership with other TFs including PU.1, RUNX1 and IRF8. In particular, PU.1 and C/EBPα have been characterised as master regulators of monopoiesis and granulopoiesis respectively with the PU.1: C/EBPα ratio instructing lineage specification (Dahl *et al*, 2003). C/EBPα interacts physically with PU.1 at DNA binding domain (DBD) locations, displacing the PU.1 co-activator C-JUN, and functionally C/EBPα blocks PU.1-induced DC development (Reddy *et al*, 2002). However, C/EBPα also acts cooperatively with PU.1, mediating chromatin accessibility at myeloid-specific enhancers (Heinz *et al*, 2010). Deletion of the *CEBPA* +37 kb enhancer in mouse models gives phenotypic effects which are identical to *CEBPA* knock-out (Avellino *et al*, 2016, Guo *et al*, 2016) whereas deletion of the -14 kb enhancer of the *SPI1* locus causes significant reduction in *SPI1* expression (Rosenbauer *et al*, 2004). Interestingly, C/EBPα can induce monocyte lineage commitment by binding at the same -14 kb enhancer, suggesting that *SPI1* expression is C/EBPα-dependent (Yeamans *et al*, 2007). More recently, the coordination of granulocyte monocyte (GM) enhancer activity by C/EBPα:PU.1 interplay during successive stages of GM-lineage differentiation (LSK, preGM, GMP, and neutrophilic granulocyte populations) has been analysed through integration of genome-wide chromatin accessibility profiles, ChIP-seq at different stages of maturation, and functional studies of preGM-KO cells from *CEBPA* $^{-/-}$ mice. These analyses suggest a more active role for C/EBPα than previously thought, acting as a pioneer factor for PU.1 at later stages of GM-lineage differentiation (Pundhir *et al*, 2018).

Furthermore, RUNX1 binds to the *CEBPA* promoter and regulates the +37 kb enhancer to favour granulopoiesis (Cooper *et al*, 2015), and *RUNX1* deletion favours monopoiesis over granulopoiesis by reducing *CEBPA* transcription (Guo *et al*, 2012). By contrast, IRF8 physically interacts with C/EBPα via its DBD, blocks C/EBPα from binding to chromatin, and inhibits neutrophil differentiation (Kurotaki *et al*, 2014). IRF8 also forms a heterodimer with PU.1 which activates an enhancer signature to promote monocyte and DC differentiation (Kurotaki *et al*, 2013). In line with these findings, *IRF8* $^{-/-}$ mice demonstrate absent monopoiesis and aberrant over-production of neutrophils (Kurotaki *et al*, 2014). Besides these direct physical interactions with partner TFs, C/EBPα can exert indirect effects, for instance it favours neutrophil differentiation by inducing microRNA-30c which in turn downregulates *NOTCH1* (Katzerke *et al*, 2013).

In addition to its pivotal role in granulocyte differentiation, C/EBPα plays a significant role in commitment to mononuclear phagocyte lineages. In this regard, studies have demonstrated that C/EBPα and PU.1 act both synergistically and antagonistically in a cell type-specific manner. For instance, C/EBPα synergises with PU.1 to reprogram differentiated B cells into macrophages (Xie *et al*, 2004). By contrast, fully-committed pre-T cells can be reprogrammed by C/EBPα and PU.1 acting antagonistically, with C/EBPα expression converting them to macrophages and PU.1 inducing them to form myeloid DCs (Laiosa *et al*, 2006).

Interestingly, the Graf group found that NOTCH1 partially inhibited and completely inhibited C/EBPα-mediated and PU.1-mediated T cell reprogramming respectively, and suggested that C/EBPα inhibits the ability of NOTCH1 to inhibit PU.1 (Laiosa *et al*, 2006). In DC progenitor cells, Carotta *et al* showed that DC differentiation cannot be induced in Flt3L cytokine conditions when PU.1 is conditionally ablated. This was partially explained by their finding that PU.1 directly controls FLT3 gene expression in DC progenitors, although enforced Flt3 expression in PU.1-deficient progenitors was unable to rescue DC differentiation (Carotta *et al*, 2010).

Regarding DC biology, the Tenen group found that C/EBPα is required for early DC differentiation both *in vitro* and *in vivo* in a transplantation model, although C/EBPα excision did not affect cytokine profiles or phagocytic function of mature DCs (Welner *et al*, 2013). In

addition, microarray of C/EBPα-deficient and wildtype DC progenitors identified gene expression changes for TFs known to be essential for DC differentiation, and ChIP-seq for C/EBPα on sorted DC progenitors confirmed direct binding of C/EBPα to associated regulatory elements (Welner *et al*, 2013). Recent single-cell studies of 5000 human $CD34^+$ cord blood cells have demonstrated that DC lineage bias starts from HSCs and is defined by transcriptional programs which correlate with IRF8:PU.1 ratios (Lee *et al*, 2017). This single-cell analysis did not elucidate the role of C/EBPα in IRF8-mediated transcriptional programs during early lineage bias, but given that C/EBPα expression mediates HSC self-renewal (Hasemann *et al*, 2014), that IRF8 interacts with C/EBPα in monocyte-DC progenitors to prevent neutrophil differentiation (Kurotaki *et al*, 2018), and the known transcriptional wiring between C/EBPα, PU.1 and IRF8 (Kurotaki *et al*, 2014), one can speculate that C/EBPα is involved in early DC lineage specification in HSCs.

In our cellular model, *CEBPA* WT-transduced cells co-express exogenous human WT *CEBPA* with endogenous bi-allelic murine WT *CEBPA*. This does not mimic physiological *CEBPA* expression but it does provide us with a comparator for transduction of human mutated *CEBPA*. Visualisation of scRNA-seq data with t-SNE reveals a distinct spatial distribution of *CEBPA* WT-transduced cells at Day 0 which corresponds with the distinct cell surface marker profile previously identified on flow cytometry. Single-cell analysis demonstrates that *CEBPA* WT-transduced cells are transcriptionally divergent from both EV and *CEBPA* N321D-transduced cells within 48 hours of transduction in the presence of nuclear Hoxb8-mediated DNA-binding. Gene expression signatures of *CEBPA* WT-transduced cells suggest ontology of granulocytes, with differentially expressed genes suggesting a GMP phenotype. Exogenous wildtype *CEBPA* is able to overcome Hoxb8 transcriptional effects and Flt3L cytokine effects, promoting both differentiation and neutrophil specification. Previous work by the Friedman group and others has shown that transcription factor ratios for C/EBPα play an important role in cell fate decisions of haematopoietic progenitors (Ma *et al*, 2014). The biological relevance of TF ratios is confirmed by our experiments which show a phenotype in *CEBPA* WT-transduced cells which is distinct from the phenotype exhibited by endogenous *CEBPA* expression.

Differential expression permits some interesting observations. *CEBPA* WT-transduced cells show expression patterns associated with lineage specification. For instance, *GFI1* is

upregulated suggesting lineage bias towards neutrophil differentiation (Karsunky *et al*, 2002, Hock *et al*, 2003). Similarly, the upregulated gene *IDH1* is known to promote myelopoiesis when overexpressed and block myeloid differentiation with knockdown in a zebrafish model (Shi *et al*, 2015). Several other genes associated with myeloid differentiation are upregulated by *CEBPA* WT at Day 2 post-transduction, including *ITGAM*, *SIRPA* and *GCA* (Roes *et al*, 2003), as well as myeloid growth factor receptor genes such as *CSF1R* (Edwards *et al*, 2019) and *CSF3R*. Upregulated genes also include genes which are antagonistic to HSPC populations, for instance transplantation of *KLF7*-transduced cells into lethally irradiated mice inhibits bone marrow reconstitution and suppresses HSCs and haematopoietic progenitor populations (Schuettpelz *et al*, 2012).

Downregulated genes include genes which are known to play important roles in DC lineage restriction, for instance *FLT3* (D'Amico *et al*, 2003, Onai *et al*, 2006) and *GATA2* (Onodera *et al*, 2016), as well as HSC-associated genes such as *SOCS2* which preserves stemness of HSCs under stress (Vitali *et al*, 2015) and *MEIS1* although not its co-factor *HOXA9* (Collins *et al*, 2014). Interestingly, *CSF3R* (upregulated) and *GATA2* (downregulated) mutations are amongst the most common co-occurring mutations in patients with *CEBPA* ^{dm} (Su *et al*, 2018, Maxson *et al*, 2016).

By contrast, the *CEBPA* N321D mutation has a remarkable effect on the WT transcriptional signature in Hoxb8-FL cells in self-renewal conditions, with only 15 upregulated and 5 downregulated genes identified on differential expression with EV-transduced cells, in comparison with 636 upregulated and 429 downregulated genes on differential expression of wildtype *CEBPA* against EV-transduced cells. This suggests that C/EBPα N321D has a dominant negative effect on exogenous wildtype C/EBPα.

In summary, *CEBPA* WT-transduced cells demonstrate a distinct transcriptional profile in the presence of Hoxb8 nuclear translocation which is characterised by upregulation of genes associated with myeloid specification and differentiation, and downregulation of genes associated with alternative lineages and HSPC states. Furthermore, the C-terminal N321D mutation is not able to promote this transcriptional signature at an early LMPP-like stage of haematopoiesis.

4.5.2 Effects of mutant C/EBPα on transcriptional wiring and myeloid differentiation, and the single-cell expression profile of *CEBPA* N321D-transduced cells in conditions which favour differentiation

Previously, we have described how Hoxb8 cells transduced with *CEBPA* N321D demonstrate distinct behaviour when differentiated in Flt3L. Specifically, *CEBPA* N321D-transduced cells demonstrate reduced expression of CD11b and SIRPα at Day 5 of differentiation, when compared to EV-transduced cells. Correlation with gene expression modules from ImmGen data suggests that Hoxb8-FL cells differentiate into both pDCs and type 2 cDCs in the presence of Flt3L. Gene set enrichment of genes differentially expressed in *CEBPA* N321D-transduced cells between Days 5 and 0 effectively points towards a pDC precursor cell ontology, both by using ASAP's enrichment tool, and by inputting upregulated genes identified by DESeq2 into an independent gene set enrichment tool such as Enrichr. In addition, the following genes were upregulated exclusively in *CEBPA* N321D-transduced cells and were associated to pDC cell ontology using Enrichr: *SLC44A1*, *TRAPPC5*, *MGAT1*, *FNDC3A*, *IFFO1*, *RUNX2*, and *RAB33B*. *RUNX2* has been shown to be highly expressed in pDCs and is required both for pDC maturation and for the expression of IRF7 which in turn regulates type I interferon production (Chopin *et al*, 2016). *RUNX2* is highly expressed in HSCs but exhibits rapidly decreased expression during myeloid differentiation at the CMP stage. Sustained *RUNX2* expression has been shown to block myeloid differentiation capacity, decrease expression of core binding factor targets *CSF1R, MPO, and CEBPD*, and decrease expression of gatekeepers of myeloid differentiation such as *CEBPA* and *GFI1* (Kuo *et al*, 2009). *SLC44A1* is upregulated in *CEBPA* N321D-transduced cells with high confidence, and encodes the surface protein CDw92, which is known to play a role in DC biology and function including the production of interleukins (Wille *et al*, 2001). Of note, Enrichr associated the upregulated genes *NOTCH1* and *ZEB2* to NK cell ontology, but these genes are known also to play a role in pDC biology. By contrast, genes shown to be downregulated exclusively between Days 5 and 0 in the *CEBPA* N321D-transduced cell line include genes (*FOSB*, *CSF3R* and *CLEC5A*) which play key roles in macrophage and granulocyte differentiation (Hong *et al*, 2011, Liu *et al*, 1996, Aoki *et al*, 2009, Batliner *et al*, 2011).

It is not possible to directly compare the gene expression profiles of *CEBPA* N321D and *CEBPA* WT-transduced cells in conditions which favour differentiation of Hoxb-FL cells, because *CEBPA* WT-transduced cells overcome Hoxb8 effects. Perhaps the most important data

subsets for differential expression at Day 5 of differentiation in Flt3L are the *CEBPA* N321D and EV-transduced single cells, because these reveal gene sets which characterise two entirely different transcriptional trajectories from an initial gene expression signature which is originally very similar between *CEBPA* N321D and EV-transduced cells in the presence of estradiol. Comparison of these two gene expression trajectories suggests that *CEBPA* N321D-transduced cells switch on a HSPC transcriptional program including *SOCS2*, and fail to switch on genes associated with macrophage and cDC programs including *SIRPA* (Gurka *et al*, 2015) and genes which encode known partner TFs of wildtype *CEBPA* such as *JUNB* and *FOS* (Rangatia *et al*, 2003, Cai *et al*, 2008, Rangatia *et al*, 2002, Hong *et al*, 2011). Notwithstanding reports by Cai *et al* that AP-1 heterodimers with C/EBPα activate the PU.1 promoter and direct monocyte lineage commitment (Cai *et al*, 2008), differential expression with DESeq2 in our experiments shows no significant differences in gene expression of either *SPI1* or *IRF8* at Day 5 between cells transduced with EV and *CEBPA* N321D. This may imply that the N321D mutation can exert its phenotypic effects by pathways which are independent of PU.1 and IRF8. However, it is important to note that our results are not directly comparable because Cai *et al* employed BM mononuclear cells which were lineage depleted and then cultured in IL-3/IL-6/SCF, rather than LMPP-like cells in cytokine conditions which promote DC differentiation.

In summary, our expression data confirms that EV-transduced Hoxb8-FL cells differentiate into mature pDCs and type 2 cDCs in the presence of Flt3L, and provides new data to suggest that *CEBPA* N321D upregulates HSPC and pDC transcriptional programs and downregulates programs associated with granulocyte, macrophage and cDC lineage restriction. Overall, this C-terminal mutation generates a pDC precursor gene expression profile when LMPP-like cells are differentiated in the presence of Flt3L. This suggests a DC-mediated mechanism which can potentially offer novel insights into the biology of C/EBPα-mutated AML, however it requires validation by further experiments. For instance, it would be interesting to further characterise this effect by taking BM mononuclear cells from a mouse model, lineage depleting them, transducing those cells with a range of C-terminal and N-terminal *CEBPA* mutations, and then culturing them in Flt3L.

In addition, single-cell analysis describes heterogeneity within the *CEBPA* N321D-transduced cells, in particular it identifies a small population of cells which continue to demonstrate a

transcriptional signature similar to immortalised Hoxb8-FL cells even in conditions where Hoxb8 is translocated to the cytoplasm. One can speculate that it is this small population which persists indefinitely in Flt3L differentiation media and which is capable of long-term engraftment *in vivo*, and therefore it would be interesting in a future experiment to employ ultra-high-throughput scRNA-seq systems such as 10x Chromium (Zhang *et al*, 2019) and capture a larger number of cells from this population to explore their gene expression profile.

5. ChIP-seq reveals the effect of *CEBPA* N321D mutation on genome-wide binding of CEBP/α and identifies direct targets for transcription

5.1 Introduction

Chromatin immunoprecipitation (ChIP) selectively enriches regions of chromatin which present specific antigen and which can be bound by a corresponding specific antibody, providing an *in vivo* snapshot of the interaction between a protein and fragments of DNA within a particular cell or tissue type. ChIP-seq builds on ChIP by employing massive parallel sequencing to annotate immunoprecipitation data onto a reference genome (Figure 5.1), and represents a powerful tool for genome-wide analysis of protein-DNA interactions (Johnson *et al*, 2007). Nevertheless, the biological interpretation of ChIP-seq data can be problematic, with a large number of binding sites having no obvious functional significance, leading some authors to hypothesise that binding events can be 'opportunistic' (Zhu *et al*, 2012). Further difficulties in interpreting ChIP-seq data are caused by evidence which suggests that TF binding patterns are commonly cell type-specific (Calero-Nieto *et al*, 2014). Ultimately, robust confirmation of TF-DNA interactions requires functional studies to establish novel transcriptional interactions and networks (Wilson *et al*, 2010).

We have found a dendritic cell phenotype on differentiating *CEBPA* N321D-transduced Hoxb8-FL cells in Flt3L, and we have employed single cell RNA-seq to characterise transcriptional profiles in EV and *CEBPA* N321D-transduced cells. Our results suggest that the *CEBPA* N321D mutation promotes an immortalised DC phenotype in conditions of differentiation in Flt3L. However, scRNA-seq does not provide any information about the mechanism of this altered gene expression profile, in particular it does not identify whether genes are direct targets of C/EBPα N321D or provide information about chromatin accessibility. Furthermore, C/EBPα is a transcription factor and exerts its biological effects by binding to enhancer and promoter regions on DNA to regulate gene expression, and there is evidence that C/EBPα can act as a pioneer factor (Di Stefano *et al*, 2016). For these reasons, it is helpful to integrate gene expression data with ChIP-seq data in our cellular model of mutant C/EBPα.

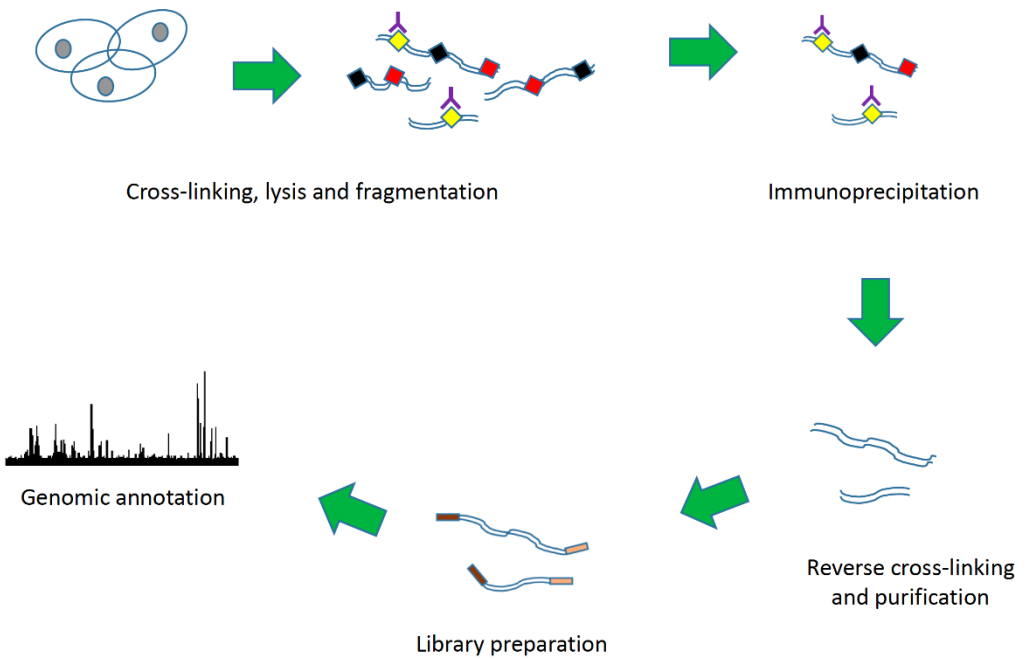


Figure 5.1 ChIP-seq combines ChIP (cross-linking of protein with DNA in chromatin followed by immunoprecipitation of DNA fragments bound by specific protein) with massive parallel sequencing to identify genome-wide binding of transcription factors

C-terminal in-frame mutations have varied effects on transactivation (Pabst *et al*, 2001), but are characterised primarily by altered DNA-binding at the bZIP region of C/EBPα. The N321D mutation is located at the 321st codon which lies within the bZIP's leucine zipper (LZ). LZs are characterised by hydrophobic leucine residues at every *d* position in an $(abcdefg)_n$ amino acid repeat, and they are configured in an α-helical structure which interdigitates with parallel residues from a partner protein at their respective LZ domains. The two dimerised TFs then bifurcate at their basic region (BR) domains to bind the major groove of DNA. Asparagine residues can specify dimer formation (Zeng *et al*, 1997), and in addition dimerisation profiles are particularly affected by amino acids at positions *a*, *e* and *g* in the LZ heptad repeat (Schuermann *et al*, 1992). *CEBPA* N321D causes asparagine to be replaced by aspartate at position *a* in the heptad repeat, and therefore it would be reasonable to hypothesise that this causes LZ reconfiguration and altered DNA-binding with associated transcriptional effects.

Genome-wide binding of C/EBPα N321D constitutes an important linking step between gene expression alterations and the altered phenotype seen in Hoxb8-FL cells transduced with the *CEBPA* mutation. This is particularly so because we know that C/EBPα is a transcription factor which dimerises and/or interacts directly with other key haematopoietic transcription factors.

For instance, C/EBPα binds to the *SPI1* promoter (Kummalue *et al*, 2003) and its -14 kb enhancer (Yeamans *et al*, 2007). In addition, evidence of TF binding can provide information on whether the genes dysregulated on gene expression analysis are direct or indirect targets of mutant C/EBPα.

ChIP-seq for C/EBPα has been previously studied in haematopoietic stem cells (Hasemann *et al*, 2014), haematopoietic progenitor and leukaemic cells (Ohlsson *et al*, 2014), adipocytes (Haakonsson *et al*, 2013), primary macrophages (Zhang *et al*, 2013), basophils and mast cells (Qi *et al*, 2013), myeloid cells including neutrophils and monocytes (Avellino *et al*, 2016) and dendritic cell progenitors (Welner *et al*, 2013). These studies have provided a number of interesting findings regarding the role of C/EBPα in haematopoiesis. Firstly, it appears that C/EBPα binds to proximal promoter and distal regulatory regions of genes implicated in myeloid differentiation as early in differentiation as the LSK cell stage of haematopoiesis (Hasemann *et al*, 2014), suggesting that C/EBPα plays a role in lineage priming before expression levels increase at the CMP to GMP junction. Secondly, C/EBPα can play a collaborative role with other transcriptional regulators such as MLL fusion proteins to promote leukaemic transformation of cells (Ohlsson *et al*, 2014). Thirdly, C/EBPα continues to modify transcriptional programs through later hierarchical cell fate decisions, for instance C/EBPα instructs specification in basophil and mast cell progenitors (Qi *et al*, 2013). Fourthly, C/EBPα-binding is tissue-specific (Di Stefano *et al*, 2014) and species-specific across vertebrate mammals (Villar *et al*, 2014). In addition, *CEBPA* expression is also tissue-specific, for instance +34 kb and +42 kb enhancers of the *CEBPA* locus are H3K27Ac-enriched only in neutrophils and monocytes, and it has been found that deletion of the +42 kb enhancer leads to tissue-specific neutropenia (Avellino *et al*, 2016). Lastly, ChIP-seq of murine CMP progenitor cells suggests that a number of genes implicated in dendritic cell differentiation are directly regulated by C/EBPα (Welner *et al*, 2013). Our own single cell RNA-seq experiments have demonstrated that *CEBPA* N321D-transduced cells display alterations in DC gene expression profiles and therefore it can be hypothesised that C/EBPα-binding to DC-related DNA may be dysregulated by mutation of *CEBPA*.

5.2 Experimental design

The objective of our experiment was to analyse the binding profile of endogenous and *CEBPA*-mutant Hoxb8-FL cells before and after differentiation for 5 days in Hoxb8 media F+e- . For this purpose, we used the following antibodies: C/EBPα (which we expect to bind to both endogenous and exogenous *CEBPA*) and FLAG (binding to exogenous *CEBPA* in the mutant phenotype only). Additionally we also studied in the same settings the distribution of H3K27Ac, which is a chromatin modification that marks regions of active chromatin. We collected nuclei from EV and *CEBPA* N321D-transduced cells derived from two distinct transfections, both in F+e+ media and after 5 days of differentiation following the withdrawal of estradiol. Cells were cross-linked with formaldehyde 1% and quenched with 2M glycine. Nuclei were subsequently purified, lysed, sonicated, pre-cleared, and then immunoprecipitated with the relevant antibody. The DNA-protein-antibody complexes were captured using agarose G beads, then eluted from beads, cross-linking was reversed, and the DNA fragments isolated by each antibody were purified.

5.3 Confirmation of ChIP enrichment by qPCR and ChIP-seq read quality

Before proceeding to the generation of a DNA library, the enrichment of the *SPI1* -14 kb upstream regulatory element (URE), known to be bound by C/EBPα, was analysed by qPCR. As a negative control, a region within mouse chromosome 1 (mChrom1) which is not bound by C/EBPα (Calero-Nieto *et al*, 2010) was also analysed. In both cases, the enrichment was compared to the DNA obtained with the IgG control, which acted as an indicator for background DNA binding.

The qPCR analysis did not evaluate genome-wide enrichment but rather enrichment at a specific locus. There was good enrichment at the *SPI1* -14 kb URE when precipitating with C/EBPα or FLAG, as compared to the mChrom1 region (Figure 5.2). The results of qPCR provided reassurance that chromatin immunoprecipitation had been effective and therefore libraries were prepared for ChIP-seq.

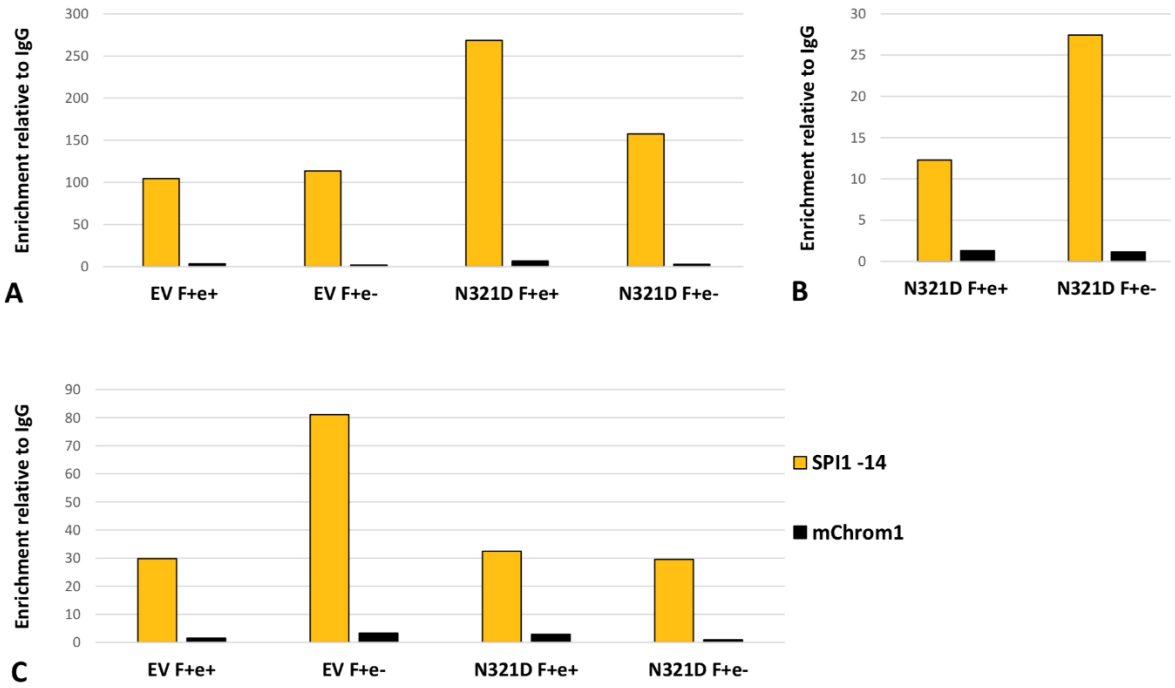


Figure 5.2 qPCR evaluation of ChIP enrichment in one of two biological replicates (A) DNA enrichment after precipitation with C/EBPα antibody is seen at higher levels at the *SPI1* -14 kb URE than at an inactive region of Chromosome 1 (Calero-Nieto *et al*, 2010). Similar results seen with (B) FLAG and (C) H3K27Ac enrichment.

Input chromatin from all four conditions was used as a control to reduce bias due to chromatin fragmentation and variable sequencing efficiency associated with different base composition. H3K27Ac antibody was used to map the distribution of histone H3 acetylated at Lysine 27 as a marker of active regions of the genome (Creyghton *et al*, 2010). C/EBPα and FLAG antibodies were employed to identify regions bound by C/EBPα and by exogenous C/EBPα N321D respectively. Representative tracks at selected enhancer regions *SPI1* -14 kb and *CEBPB* +65 kb are shown as evidence for ChIP-seq quality (Figure 5.3):

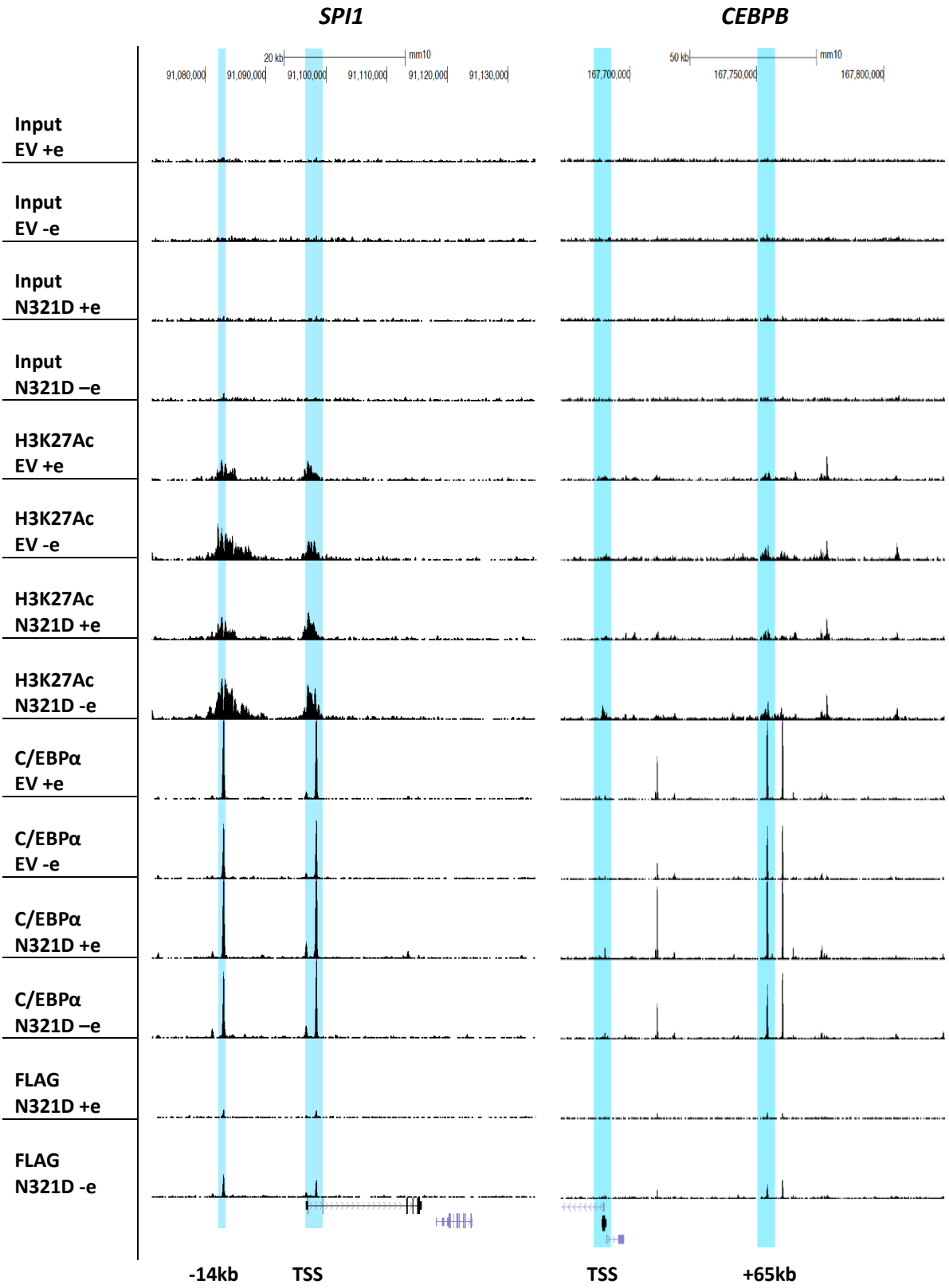


Figure 5.3 UCSC Genome Browser screenshots of Input, H3K27Ac, C/EBPα and FLAG binding at selected enhancers *SPI1* -14 kb and *CEBPB* +65 kb

5.4 Binding events for exogenous mutant C/EBPα, but not endogenous C/EBPα, are markedly increased during differentiation of mutant cells in Flt3L

Figure 5.4 represents DNA regions of increased enrichment at Days 5 and 0 for C/EBPα and FLAG antibodies. Interestingly, marked alteration in binding was only seen in the FLAG antibody for both ChIP-seq replicate experiments.

Figures 5.4 (A) and (B) represent replicate experiments for enrichment by C/EBPα of DNA taken from EV cells, comparing enrichment for DNA regions at Day 5 and at Day 0. Figures 5.4 (C) and (D) represent the same for DNA taken from *CEBPA* N321D-transduced cells. Figures 5.4 (E) and (F) represent replicate experiments for enrichment by FLAG of DNA taken from *CEBPA* N321D-transduced cells. Enrichment corresponds to the log2 of normalised read counts at C/EBPα and FLAG peaks respectively.

The FLAG sequence was inserted into the N321D construct only, and therefore the FLAG antibody enriches only for DNA regions bound by C/EBPα N321D and not for regions bound by endogenous C/EBPα.

For C/EBPα, there is no obvious quantitative change in global enrichment of DNA between Day 5 and Day 0 in either EV or *CEBPA* N321D-transduced cells, although specific DNA regions are bound differentially between the days and between the conditions (Figures 5.5, 5.6 and 5.7). In contrast with the bi-modal distribution of global read counts at C/EBPα peaks between Days 5 and 0, there were relatively few regions with increased read counts at FLAG peaks at Day 0 in estradiol, compared to Day 5 after differentiation in Flt3L. It is possible that at Day 0 the C/EBPα-bound complexes are relatively stable and inaccessible by the mutant, but upon differentiation there is a higher requirement for C/EBPα and its partners for protein synthesis, and at this point the mutant TF binds to DNA and exerts a dominant negative effect on the wildtype TF. One can also speculate that genes for lineage specification and differentiation are preferentially bound by the mutant TF. Interestingly, the wildtype C/EBPα binds equally avidly during progenitor and differentiated cell states, whereas the mutant C/EBPα binds more avidly during differentiated cell states, supporting the hypothesis that Leukaemia initiating cells in C/EBPα-mutated AML are differentiated cells rather than HSPCs.

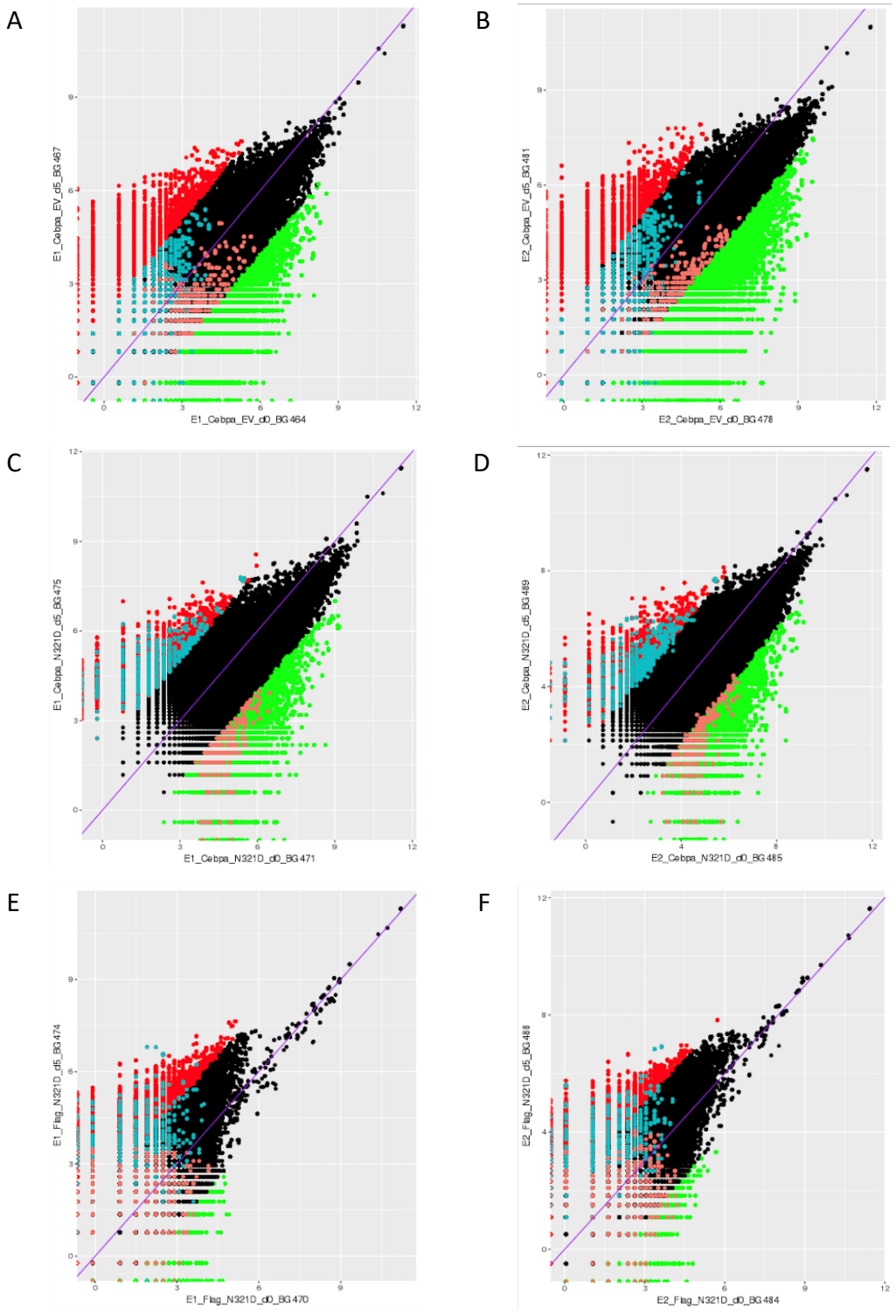
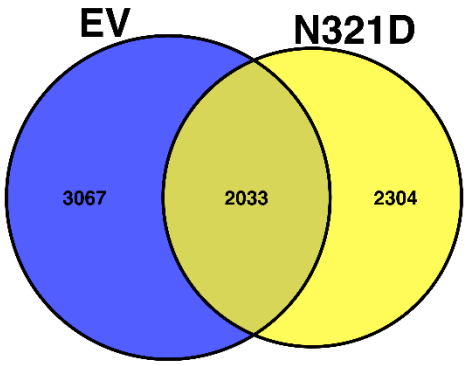


Figure 5.4 Read count comparison for two biological replicates between Day 0 and after differentiation in Flt3L for 5 days, for (A) and (B) C/EBPα antibody in EV cells, (C) and (D) C/EBPα antibody in mutant cells, and (E) and (F) FLAG antibody in mutant cells only. Log2 graph with red dots representing DNA regions with ≥ 4-fold increase in read count and green dots representing ≥ 4-fold decrease in read count when comparing Days 5 and 0. Line corresponding to direct correlation (no change) is depicted. Blue and pink dots represent regions enriched and depleted respectively for N321D-transduced cells when compared to EV cells.

Exogenous C/EBPα WT binding has not been analysed by ChIP-seq, but the altered phenotype and gene expression profile suggest that a higher ratio of C/EBPα in *CEBPA* WT-transduced cells may act to open up chromatin in cells even before they are cultured in conditions which favour differentiation. This would be consistent with recent reports by Thomas Graf and colleagues that pulsed *CEBPA* expression in B cells can open up chromatin and poise them for reprogramming into induced pluripotent stem cells (Di Stefano *et al*, 2014).

Binding events from both biological replicates in EV and *CEBPA*-transduced cells were then compared to find DNA regions unique to each cell line which demonstrated four-fold increased and decreased read counts for C/EBPα-binding between Days 5 and 0 of differentiation (Figures 5.5, 5.6).

A



B

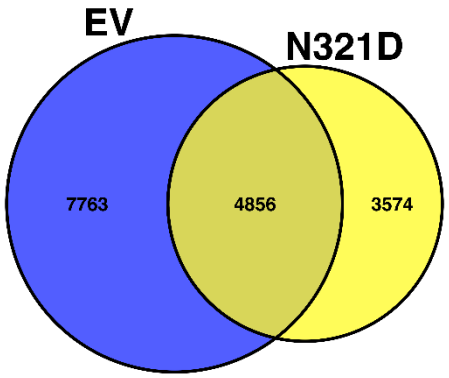


Figure 5.5 Regions of DNA which demonstrate four-fold increased (A) and decreased (B) read counts for C/EBPα-binding in first biological replicate between Days 5 and 0 of differentiation in Flt3L

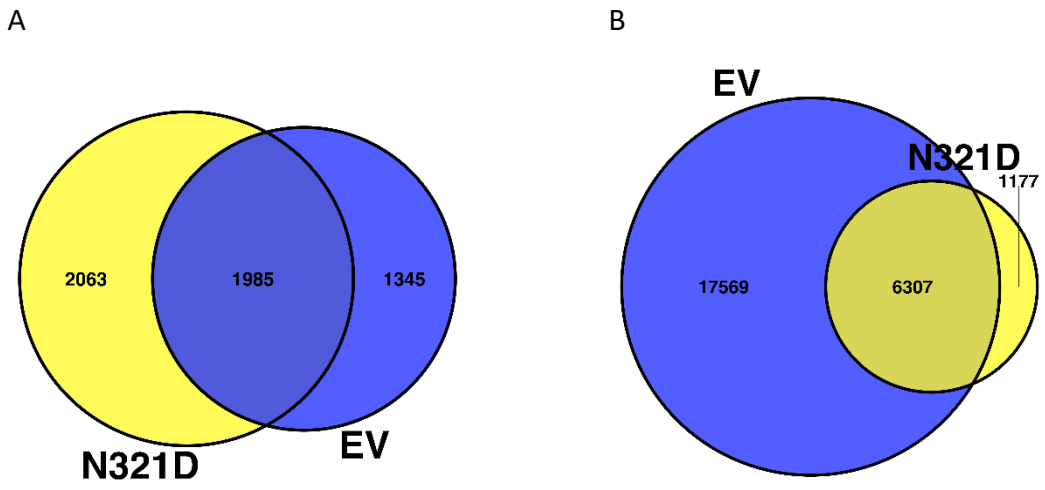


Figure 5.6 Regions of DNA which demonstrate four-fold increased (A) and decreased (B) read counts for C/EBPα-binding in second biological replicate between Days 5 and 0 of differentiation in Flt3L

Results from the two biological replicates were then intersected to identify a final list of high confidence regions where C/EBPα binding is upregulated (Figure 5.7 A) or downregulated (Figure 5.7 B) during differentiation in Flt3L:

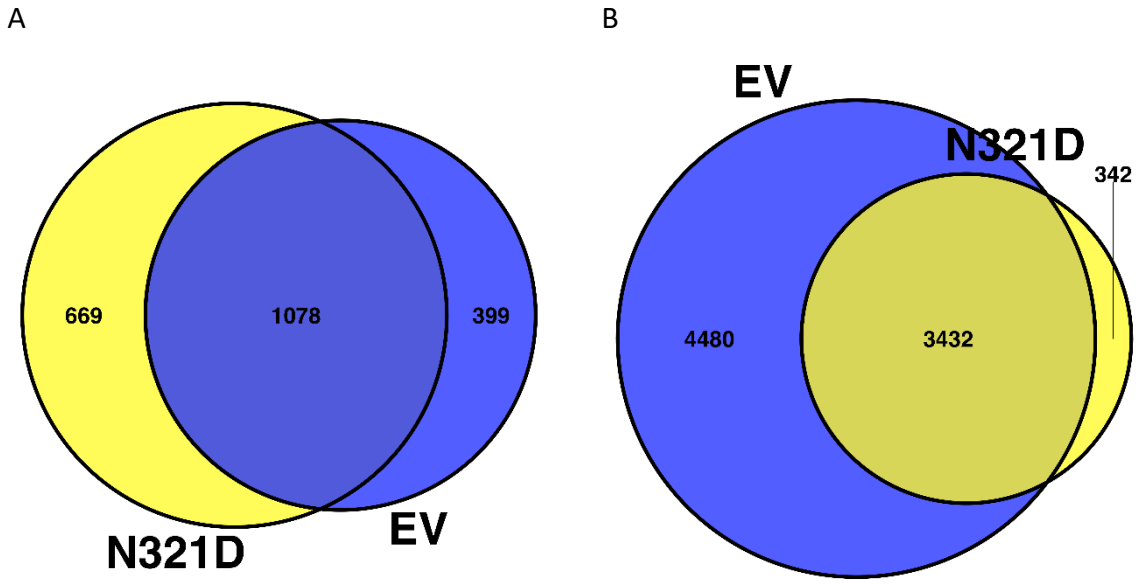


Figure 5.7 High confidence regions of DNA which demonstrate four-fold increased (A) and decreased (B) read counts for C/EBPα-binding after intersection of both biological replicates

5.5 Intersection of differential gene expression and C/EBPα binding events

Following the observation that there is a quantitative shift in FLAG genome-wide binding events but not C/EBPα binding events between Days 0 and 5 (Figure 5.4), and bearing in mind that there is a phenotypic difference between these two cell lines, it was decided to carry out a qualitative analysis of regions bound by C/EBPα.

We examined genome-wide binding profiles of EV and *CEBPA* N321D-transduced cells at Day 5 of differentiation, by defining genes which were associated with regions bound by C/EBPα in EV and *CEBPA* N321D-transduced cells, and by FLAG in *CEBPA* N321D-transduced cells only. We identified 4752 genes with an associated C/EBPα peak in both experimental replicates for EV cells, and 8843 genes with a C/EBPα peak in both replicates for *CEBPA* N321D-transduced cells. This represents a significant difference in number of genes bound by C/EBPα in EV and *CEBPA* N321D-transduced cells at Day 5 of differentiation in Flt3L, and it is possible that either the increased expression ratio of *CEBPA* in *CEBPA* N321D-transduced cells or else binding effects specific to C/EBPα N321D, for instance conformational changes in the TF, are responsible for the quantitative difference. The 8843 genes associated with a C/EBPα peak included binding events for both endogenous murine C/EBPα and for exogenous human C/EBPα N321D, and therefore we also identified 2747 genes with an associated FLAG peak in both experimental replicates for *CEBPA* N321D-transduced cells. These 2747 genes were specifically bound by exogenous C/EBPα N321D.

This data was then correlated with gene expression data from our scRNA-seq experiment. During differentiation of EV cells between Days 0 and 5, 1400 genes were upregulated and 1340 genes were downregulated. By contrast, during differentiation of *CEBPA* N321D-transduced cells over 5 days, 594 genes were upregulated and 418 genes were downregulated (Figure 5.8). In addition, differential expression between both cell lines at Day 5 gives 224 upregulated and 633 downregulated genes in *CEBPA* N321D-transduced cells when compared to EV cells at Day 5 (Figure 5.9).

This data can then be interrogated by examining the 1400 genes which are upregulated during differentiation of EV cells, identifying those genes which have an associated C/EBPα peak in both ChIP-seq replicate experiments at Day 5 of differentiation (Appendix, Table 11), and then exploring the binding patterns of those same 653 genes in *CEBPA* N321D-transduced cells. This strategy looks at differential expression in EV cells between Days 5 and 0, and then correlates this with ChIP-seq binding profiles (see Figure 5.8).

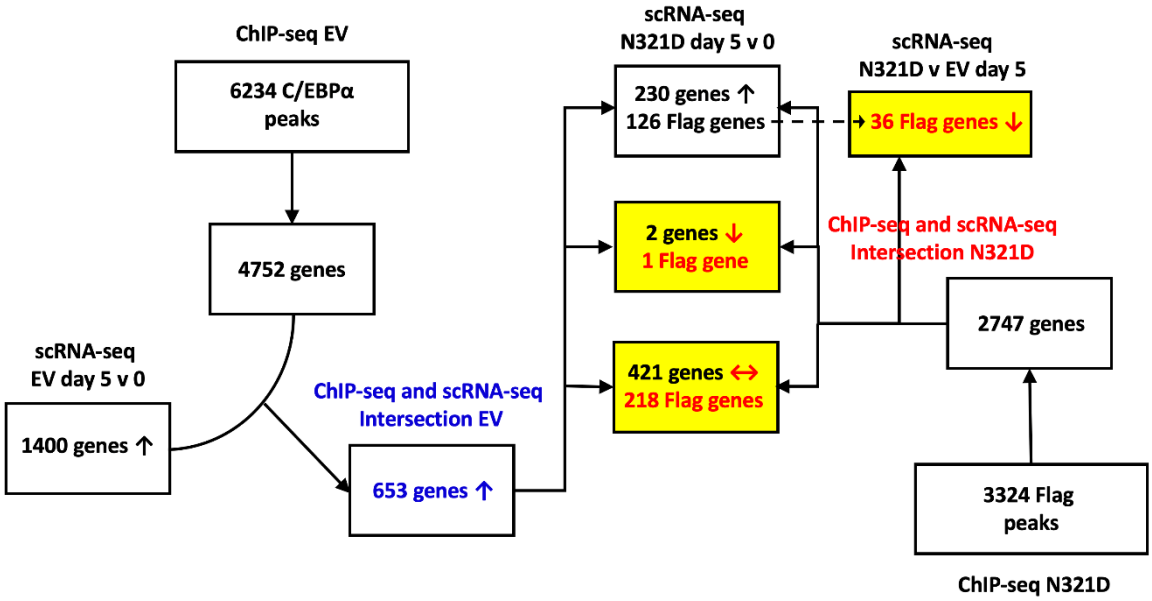


Figure 5.8 Schematic of first strategy to analyse qualitative variations in genome-wide binding profiles between EV and *CEBPA* N321D-transduced cells

From the 653 genes identified as upregulated and having an associated C/EBPα peak in EV cells, 230 genes are also upregulated in *CEBPA* N321D-transduced cells during differentiation (Appendix, Table 12), and 126 are bound by the FLAG antibody at Day 5 (Appendix, Table 13). Interestingly, 36 of these 126 genes are downregulated in *CEBPA* N321D-transduced cells when compared to EV cells at Day 5. These 36 genes (Appendix, Table 14) are therefore bound by exogenous C/EBPα N321D, upregulated during the differentiation process in *CEBPA* N321D-transduced cells, but their upregulation is impaired in comparison with EV cells. This means that although these genes are upregulated during differentiation of *CEBPA* N321D-transduced cells, exogenous C/EBPα N321D binding events seem to have a dominant negative effect on expression of these genes when compared to EV-transduced cells.

From the 653 genes, 2 genes are downregulated in *CEBPA* N321D-transduced cells during differentiation and 1 of those 2 genes is also bound by the FLAG antibody.

421 of the 653 genes do not significantly change expression levels during differentiation of *CEBPA* N321D-transduced cells, including 218 genes which are bound by the FLAG antibody at Day 5 in *CEBPA* N321D-transduced cells. These 218 genes confirm the general observation that ChIP-seq binding events do not necessarily alter expression profiles or biological function in each and every cellular context (Calero-Nieto *et al*, 2014).

In addition to upregulated genes during differentiation of EV cells in Flt3L which have an associated C/EBPα peak in EV cells, there are 1340 downregulated genes identified on differential expression between EV cells at Days 5 and 0, of which only 3 genes (*BC017158*, *AI314180* and *AI987944*) have an associated C/EBPα peak in EV cells. This suggests that there is little evidence of a correlation between C/EBPα-binding and downregulation of gene expression during differentiation of EV cells and therefore further exploration of this data set was not carried out.

5.5.1 Immunomodulatory and DC genes including *NOTCH2* are bound by mutant C/EBPα and demonstrate reduced upregulation in *CEBPA* N321D-transduced cells during differentiation in Flt3L

These 36 genes (see Figure 5.8) represent genes which are bound by both endogenous C/EBPα and exogenous C/EBPα N321D, upregulated during differentiation in both EV and *CEBPA* N321D-transduced cells, but their upregulation is impaired in comparison with EV cells. They include immunoregulatory genes involved in DC differentiation and function such as *AXL*, which acts as a canonical marker of a discrete population of pDCs which do not produce type I IFN (Zhang *et al*, 2017) and which is upregulated by IFN-α during differentiation of human DCs (Scutera *et al*, 2009). Expression of *CD300a*, another of these 36 genes, is expressed by a wide range of immunomodulatory cells including NK cells, T and B cells, monocytes and dendritic cells (Clark *et al*, 2009). *CD300a* is downregulated by exogenous IFN-α, and *CD300a* in turn regulates IFN-α and TNF-α secretion by human pDCs (Ju *et al*, 2008).

ITGAM encodes CD11b, which is highly expressed in both monocytes and dendritic cells, and *ITGAM* has been found to regulate lipopolysaccharide pathways in myeloid DCs but not in

macrophages. In addition, *ITGAM*/CD11b promote TLR4 endocytosis and TRIF-mediated signalling in myeloid DCs (Ling *et al*, 2014). CD11b provided the first evidence of a DC phenotype in our initial flow cytometry experiments, and shows the important role that can be played by genes which are upregulated during differentiation of *CEBPA* N321D-transduced cells but have impaired upregulation in comparison with EV cells.

Genes including *TLR2* and *CTSB* are implicated in tolerogenic DC mechanisms. *TLR2* upregulation in tolerogenic DCs results in a reduced proinflammatory immune program (Chamorro *et al*, 2009) and *TLR2* activation promotes tumour DC dysfunction whereas *TLR2* blockade improves the efficacy of immunotherapy (Tang *et al*, 2015). *CTSB* is involved in apoptosis and lysosomal activation in pDCs (Robbins *et al*, 2008) and in MHC class II antigen presentation in tolerogenic DCs (Schinnerling *et al*, 2015).

Other genes which demonstrate reduced upregulation in *CEBPA* N321D-transduced cells have not been directly associated with DC differentiation or function, but are involved in immunoregulatory mechanisms. For instance, *DPEP2* mediates inflammatory effects through conversion of leukotriene D4 to leukotriene E4 (Toda *et al*, 2013). Similarly, *IFITM3* is an antiviral gene stimulated by type 1 IFN, and expression of *IFITM3* is increased by A/Bris/59/07 infection (Cao *et al*, 2012).

From the perspective of dendritic cell function, perhaps the most significant gene bound by FLAG which has impaired upregulation in *CEBPA* N321D-transduced cells in comparison with EV-transduced cells is *NOTCH2*. The NOTCH family of transmembrane receptor TFs mediates an intercellular signalling pathway which regulates decisions between adjacent cells. Each family member comprises a heterodimeric receptor with extracellular subunits on the cell surface engaging ligands presented by *trans*- or opposing cells, and intracellular subunits mediating gene expression patterns associated with cell fate decision and differentiation. In addition, NOTCH mediates interaction between bone marrow stroma (BMS) and haematopoietic progenitor cells. The NOTCH ligand Jagged-1 is highly expressed in BMS and stimulates accumulation of DC precursors while preventing terminal differentiation of DCs, whereas the Delta-1 ligand is highly expressed in spleen stroma and promotes terminal differentiation (Cheng *et al*, 2007). Interestingly, deletion of the NOTCH2 receptor causes a tissue-specific reduction of DC populations in the spleen (Lewis *et al*, 2011). This suggests that

the effect of NOTCH receptor/transcription factors on DC differentiation is complex and dependent on cellular or spatial context. In addition, NOTCH signalling mediates immunomodulatory mechanisms in cancer, for instance it mediates functionally altered myeloid cells and NOTCH inhibition in DCs can cause abnormal myeloid cell differentiation which is implicated in oncogenesis (Cheng *et al*, 2014). The observation that *NOTCH2* is bound by FLAG and shows impaired gene expression in *CEBPA* N321D-transduced cells suggests that C/EBPα N321D may directly interact with *NOTCH2 cis*-regulatory elements and interfere with DC differentiation.

Taken together, these data demonstrate that genes directly bound by C/EBPα N321D which show decreased expression during differentiation of *CEBPA* N321D-transduced cells include genes specifically associated with both pDC and cDC differentiation and function, as well as more generic immunomodulatory mechanisms including tolerogenicity and inflammation.

5.5.2 Loci for important partner TFs (PU.1, RUNX1, IRF8) are bound by exogenous mutant C/EBPα but have no altered gene expression

126 genes are bound by endogenous C/EBPα, upregulated during differentiation in both EV and *CEBPA* N321D-transduced cells, and bound by exogenous C/EBPα N321D. They include the 36 genes identified in 5.5.1 above, but also 90 genes which are bound by exogenous C/EBPα N321D and have no significant difference in expression levels between EV and *CEBPA* N321D-transduced cells at Day 5 of differentiation. It is important to note that these genes are specifically bound by the mutant C/EBPα N321D, because they encode important partner TFs for C/EBPα which are known to play a role in dendritic cell function and/or differentiation such as IRF8 (Kurotaki *et al*, 2014, Sichien *et al*, 2016), PU.1 (Carotta *et al*, 2010, Merad, 2010) and RUNX1 (Guo *et al*, 2012, Satpathy *et al*, 2014). It is well-established that TFs act in combinatorial networks (Wilson *et al*, 2010) and it is possible that C/EBPα N321D binding may produce secondary indirect effects even if expression of *SPI1* is not significantly altered at Day 5 of differentiation. Previous work by Miller *et al* has shown that mutations in bZIP TFs including C/EBPα alter conformation and binding specificity at the level of amino acid residues (Miller *et al*, 2003), and more recent work has demonstrated that mutations can be engineered to alter the dimerisation profiles of bZIP TFs (Pogenberg *et al*, 2014). In addition, Thomas Graf and colleagues have recently shown that TFs can drive topological genome reorganisation before changes in gene expression (Stadhouders *et al*, 2018). Therefore one

can hypothesise that mutant C/EBPα may alter conformational interactions with partner TFs even if it does not change the expression level, and this can in turn then change the combinatorial effect of C/EBPα with the partner TF. It would be interesting to employ crystallography or chromosome conformation capture to examine the conformational effects of the N321D mutation but this is outside of the scope of this study. However, it is important to recognise that the interplay between TFs and three-dimensional genome topology can drive cell fate decisions (Stadhouders *et al*, 2019).

5.5.3 *TGFBI* is bound by mutant C/EBPα, upregulated during differentiation in EV cells, but downregulated during differentiation of *CEBPA* N321D-transduced cells

From the 653 genes which are bound by endogenous C/EBPα and upregulated during differentiation in EV cells, 2 genes (*FCGR3* and *TGFBI*) are downregulated in *CEBPA* N321D-transduced cells during differentiation and 1 of those 2 genes is also bound by the FLAG antibody. This single gene *TGFBI* represents a reversal of expression profile between *CEBPA* N321D-transduced cells and EV cells which is directly associated with binding of mutant C/EBPα N321D.

The role of *TGFBI* in the pathogenesis of corneal dystrophy has been elucidated in several studies (Venkatraman *et al*, 2019), and altered expression of *TGFBI* has been implicated in non-small cell lung cancer (Yan *et al*, 2018), glioma cells (Guo *et al*, 2018, Pan *et al*, 2018), and other cancers (Han *et al*, 2015, Zhu *et al*, 2015, Ozawa *et al*, 2016, Lauden *et al*, 2014). *TGFBI* has been characterised as a regulator of HSC specification in a zebrafish model (Charbord *et al*, 2014), and plays a regulatory role in human haematopoiesis that includes cell-intrinsic effects but which also involves interplay between both human HSPCs and bone marrow mesenchymal cells (Klamer *et al*, 2018). Specifically, HSPC maintenance was enhanced by reduced expression of *TGFBI*. In addition, *TGFBI* has been shown in microarray analysis to have increased expression in a $CD14^{+}$ myeloid subset of human dendritic cells (Ahn *et al*, 2002), and to have decreased expression in DCs where maturation was attenuated by regulatory T cells (Mavin *et al*, 2017). It can be hypothesised that reduced expression of *TGFBI* in *CEBPA* N321D-transduced cells associated with exogenous C/EBPα N321D binding mediates HSPC maintenance and contributes to a differentiation block in DC maturation.

5.5.4 Dominant negative effect of C/EBPα N321D affects genes implicated in DC cell fate and function, as well as genes involved in cell cycle control

From the 653 genes which were bound by endogenous C/EBPα and upregulated during differentiation in EV cells, 421 genes showed no significant differential expression in *CEBPA* N321D-transduced cells during differentiation over 5 days, and 218 of these genes were bound by the FLAG antibody.

These 218 genes represent the genes directly bound by C/EBPα N321D with a failure to follow the EV model of gene upregulation during differentiation in Flt3L. These genes are the most likely to be subject to a dominant negative effect from the N321D mutation. They include genes which are confirmed by other intersections in this chapter (see 5.6.1 below) and which play a role in DC lineage specification and/or function, including *COTL1*, *DAB2*, *FOS*, *IRAK2*, *JAK2* and *SIRPA*.

The downregulation of *FOS* in *CEBPA* N321D-transduced cells is particularly interesting because the AP-1 proteins are known to act as dimerisation partners of C/EBPα. Previous work has shown that C/EBPα:C/EBPα homodimers promote granulopoiesis whereas C/EBPα:Fos heterodimers induce monocyte lineage commitment, and that the interaction between C/EBPα and AP-1 TFs plays a significant role in cell fate decision-making at the CMP to GMP junction (Cai *et al*, 2008).

Zhong *et al* have shown that JAK2 is essential for bone marrow derived dendritic cells (BMDCs) development and maturation, and JAK2 deficiency attenuates innate immune response though not adaptive immune response (Zhong *et al*, 2010). This study did not specifically examine cDC or pDC populations, but it suggests that downregulation of *JAK2* in *CEBPA* N321D-transduced cells may cause a maturation defect in the DC lineage.

COTL1 alongside *AXL* acts as a canonical marker of a novel discrete subpopulation of pDCs which do not produce type I IFN, but it is not highly expressed in the majority of pDCs which secrete IFN-α and are $CD5^-$ $CD81^-$ (Zhang *et al*, 2017).

IRAK2 plays a role in interferon production by pDCs, with *IRAK2*-deficient pDCs demonstrating increased nuclear translocation of IRF7 which acts as a key TF for interferon gene transcription (Wang *et al*, 2011).

DAB2 is induced during differentiation of murine bone marrow derived DCs and human monocyte-derived DCs and appears to act as a negative regulator of DC immunogenicity and T cell activation (Ahmed *et al*, 2015). Downregulation of *DAB2* in *CEBPA* N321D-transduced cells may therefore promote DC migration, antigen uptake, T-cell responses and secretion of cytokines.

In addition to genes which play a role in cell fate decisions and dendritic cell function, this subset of genes also includes three genes which play a role in cell cycle control and therefore may be relevant for the pre-leukaemic effects of the *CEBPA* N321D mutation. Firstly, *CXCR4* has been shown to be associated with cell cycle arrest in leukaemic cells (Burger *et al*, 2007) and furthermore wildtype C/EBPα directly binds to the *CXCR4* promoter to activate transcription (Kuo *et al*, 2014). In addition, it has been shown that an N-terminal mutation of *CEBPA* impairs *CXCR4* expression (Kuo *et al*, 2014). Our own observation that *CXCR4* expression is upregulated in EV and not *CEBPA* N321D-transduced cells suggests that binding by the C-terminal mutant C/EBPα N321D may similarly impair expression of *CXCR4* in the Hoxb8-FL cell line.

KLF4 encodes one of the Yamanaka factors (Takahashi *et al*, 2006) and is a second gene which has been implicated in cell cycle control, mediating p53-activation of the cyclin-dependent kinase inhibitor p21 (Zhang *et al*, 2000). In addition, *KLF4* is a target gene of PU.1 and acts as a critical regulator of monopoiesis, promoting monocyte differentiation (Feinberg *et al*, 2007). Failed upregulation of *KLF4* in *CEBPA* N321D-transduced cells may impair both monopoiesis and cell cycle control in mutant cells.

Finally, E2F2 has been shown to repress the transcription of cell cycle genes and promote a quiescent G0 state (Infante *et al*, 2008). In Drosophila, E2F2 antagonises E2F1 and acts to suppress both transcription of E2F-regulated genes and DNA synthesis (Frolov *et al*, 2001), and E2F2 has been shown to act as a tumour suppressor in specific cellular contexts (Opavsky *et al*, 2007). Decreased expression of *E2F2* is particularly interesting because it is well-

established that C/EBPα mediates cell growth arrest through E2F-mediated cell cycle mechanisms (Slomiany *et al*, 2000, Kowenz-Leutz *et al*, 2016).

5.6 Intersection of mutant C/EBPα (FLAG) binding events and differential gene expression between cell lines at Day 5

The second strategy is to look first at the genome-wide binding profile of the FLAG antibody in *CEBPA* N321D-transduced cells, and then correlate this profile with gene expression data. There are 2747 genes with an associated FLAG peak at Day 5 in both replicates of the ChIP-seq experiments. 160 of these 2747 genes are downregulated (Appendix, Table 15), 49 genes are upregulated (Appendix, Table 15), and 2539 out of the 2747 genes do not significantly change expression profile in *CEBPA* N321D-transduced cells when compared to EV cells at Day 5. Interestingly, it appears that most of the FLAG binding events are not associated with altered transcriptional status of directly bound genes, and that the dramatic phenotypic difference between the two cell lines is achieved with relatively few downregulated and even fewer upregulated genes.

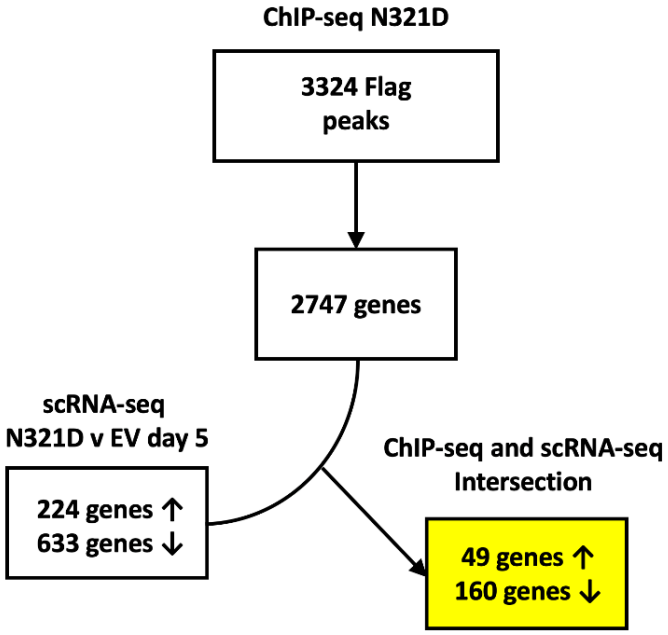


Figure 5.9 Schematic of second strategy to analyse qualitative variations in genome-wide binding profiles between EV and *CEBPA* N321D-transduced cells

5.6.1 Regulatory elements for genes involved in myeloid differentiation and DC specification are bound by mutant C/EBPα and show reduced expression in mutant cells at Day 5 of differentiation in Flt3L

These 160 genes represent genes which are directly bound by exogenous mutant C/EBPα and which are upregulated in EV-transduced cells or downregulated in *CEBPA* N321D-transduced cells on differential expression between these cells at Day 5 of culture in media with Flt3L. There is significant overlap with the subset of 36 genes which are bound by exogenous FLAG, upregulated during differentiation in *CEBPA* N321D-transduced cells, but demonstrate impaired upregulation in comparison with EV cells, with shared genes between the subsets including *AXL*, *CD300a*, *CTSB*, *DPEP2*, *ITGAM*, *TLR2* and *NOTCH2*.

Nevertheless, there are some additional genes in this data set which are relevant for DC biological processes, and which may be directly regulated by the *CEBPA* N321D mutation in the Hoxb8-FL cell line. These genes include well-established genes involved in myeloid differentiation pathways such as *CSF3R*, *JAK2* and *FOS*, as well as other genes which have been shown to play roles in DC differentiation and function.

The role of *JAK2* and *FOS* has been described in 5.5.4 above. *CSF3R* is downregulated on differential expression of *CEBPA* N321D versus EV-transduced cells at Day 5 of culture in media with Flt3L. *CSF3R* encodes the G-CSF receptor which plays an important role in neutrophil ontogeny. Increased expression of *CSF3R* has also been demonstrated in $CD11b^{+}$ cDCs (Robbins *et al*, 2008), and downregulation of *CSF3R* in *CEBPA* N321D-transduced cells compared with EV-transduced cells suggests an altered DC gene expression program. One can speculate that the upregulation of *CSF3R* in the presence of Flt3L contributes to the cDC phenotype seen in EV-transduced cells, and that failed upregulation in *CEBPA* N321D-transduced cells allows the pDC progenitor-like phenotype to persist.

Other genes upregulated in EV-transduced cells include genes associated with identification of DCs or specific DC subsets (*CLEC10A*, *CLEC4A1*, *COTL1*, *CD9*) and genes associated with DC function (*DAB2*, *IRAK2*, *MARCH1* and *CD9* again). The roles of *COTL1*, *DAB2* and *IRAK2* have been previously mentioned in 5.5.4 above.

CLEC10A is a specific marker for human $CD11c^+$ dendritic cells which are homologous to murine $CD8^-$ $CD11b^+$ $SIRP\alpha^+$ type 2 cDCs (Heger *et al*, 2018). Increased expression of *CLEC10A*, and of the *SIRPA* gene, on differential expression of EV as compared to *CEBPA* N321D-transduced cells therefore corresponds with flow cytometry data which suggests that EV cells differentiate into type 2 cDCs in Flt3L and that *CEBPA* N321D-transduction causes a differentiation block in this process.

CLEC4A1 encodes Dendritic cell inhibitory factor 4 (DCIR4) and its expression is significantly diminished by differentiation of inflammatory monocytes into DCs (Kameda *et al*, 2016).

CD9 is downregulated on differential expression between *CEBPA* N321D and EV-transduced cells. Importantly, pDCs have reduced surface expression of MHC class II molecules in comparison with cDCs, and pDCs do not express *CD9* (Asselin-Paturel *et al*, 2001, Mittelbrunn *et al*, 2009). The downregulation of *CD9* in *CEBPA* N321D-transduced cells may reflect a differentiation shift in mutant cells from cDC to pDC phenotype.

In addition to discriminating between cDC and pDC subsets, *CD9* also plays a role in DC function. It has been shown to facilitate superior antigen presentation ability in murine DCs compared to other APCs by playing a role in physically tethering dissimilar MHC class II molecules (Unternaehrer *et al*, 2007). *CD9* deletion limits MHC II surface expression and significantly impairs $CD4^+$ T cell activation in monocyte-derived DCs (Rocha-Perugini *et al*, 2017).

DAB2 and *MARCH1* are involved in antigen presentation and T-cell activation mediated by dendritic cells. MARCH1-mediated MHC II ubiquitination in DCs regulates MHC class II surface expression and immunogenic T-cell activation, but also promotes DC selection of regulatory T cells (Oh *et al*, 2013).

In summary, our results suggest that C/EBPα N321D binds to regulatory elements for genes which play significant roles in myeloid differentiation, DC specification and DC function, blocking recruitment of relevant co-activators and general transcription factors.

5.6.2 Mutant C/EBPα binds to HSPC genes including *MYB* and increases their expression in Flt3L differentiation media

These 49 genes represent genes which are directly bound by exogenous mutant C/EBPα and which are downregulated in EV-transduced cells (or upregulated in *CEBPA* N321D-transduced cells) on differential expression between these cells at Day 5 of differentiation in Flt3L.

Interestingly, these 49 genes include genes which are important for maintenance of haematopoietic progenitor populations. For instance, *SDC1* transcripts are upregulated in murine MPP cells, in contrast to *SDC2* transcripts which are enriched in LT-HSCs (Forsberg *et al*, 2005). *CBX2* is another gene in this subset, and is required for both haematopoietic and progenitor self-renewal with *CBX2* shRNA knock-down resulting in reduced plating potential (van den Boom *et al*, 2013). Perhaps the most interesting gene upregulated by *CEBPA* N321D-transduced cells is *MYB*. Primitive haematopoiesis does not require *MYB* (Tober *et al*, 2008), but foetal liver haematopoiesis is MYB-dependent (Mucenski *et al*, 1991) and LT-HSCs express the highest levels of *MYB* with progressive decrease in *MYB* expression during step-wise differentiation to MPP cells and both granulocytes and monocytes (Lieu *et al*, 2012).

Other genes which play a role in DC maturation include *SOCS2* in human monocyte-derived DCs (Hu *et al*, 2009) and *IL6ST* which encodes a signal transducer for IL-6, which in turn blocks DC maturation through STAT3 activation in a mouse model (Park *et al*, 2004).

5.7 Intersection of chromatin accessibility profile and C/EBPα binding events during differentiation in Flt3L

The third strategy to analyse ChIP-seq data is to describe H3K27Ac coverage in C/EBPα-bound regions in both EV and *CEBPA* N321D-transduced cells to examine qualitative changes in chromatin accessibility in the two conditions. This can be achieved firstly by defining C/EBPα-bound regions in each cell line, and then by intersecting this data with histone acetylation data.

C/EBPα-bound regions can be defined as follows:

(i) identify regions which are enriched four-fold for C/EBPα (Figure 5.4) when comparing each cell line at Day 5 versus Day 0, identifying those regions which are differentially bound during differentiation in each cell line;

(ii) intersect results from the two ChIP-seq experiments in each cell line to identify high confidence peaks for C/EBPα enrichment during differentiation;

(iii) intersect the high confidence regions from the two cell lines to identify 3 sets of regions (Figure 5.10) which are highly enriched for C/EBPα binding specifically in EV cells (399 regions) or in *CEBPA* N321D-transduced cells (669 regions) or which are enriched in both cell lines (1078 regions).

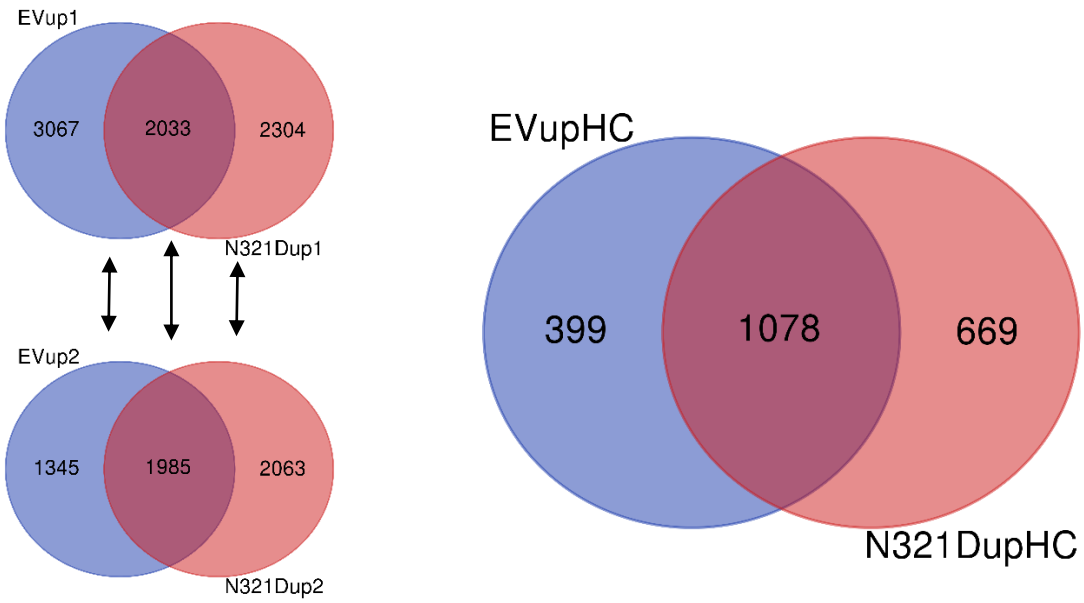


Figure 5.10 C/EBPα-bound regions enriched four-fold and intersecting two ChIP-seq experiments to establish high confidence peaks

The peak distribution of the regions in Figure 5.10, and the total peak distribution over all C/EBPα binding events can be analysed to determine whether the regions enriched by C/EBPα binding are more likely to be associated with promoters or enhancers (Table 5.1).

Table 5.1 Peak distribution of regions enriched four-fold by C/EBPα-binding

	Promoter	Intragenic	Intergenic
399 up (EV) peaks	7	217	175
669 up (N321D) peaks	11	321	337
1078 up (shared) peaks	17	436	625
All peaks	15317	53225	52933

The peak distribution listed in Table 5.1 suggests that, in comparison with all C/EBPα binding events, high confidence regions enriched four-fold for C/EBPα are more likely to be associated with intragenic and intergenic enhancers than with promoters.

Chromatin accessibility in the C/EBPα-bound regions can then be analysed as follows:

(iv) examine acetylation coverage with H3K27Ac-binding of the 3 sets of regions (Figure 5.11);

(v) identify regions which are highly acetylated (two-fold enrichment of H3K27Ac) in C/EBPα-enriched regions;

(vi) map the identified regions to genes.

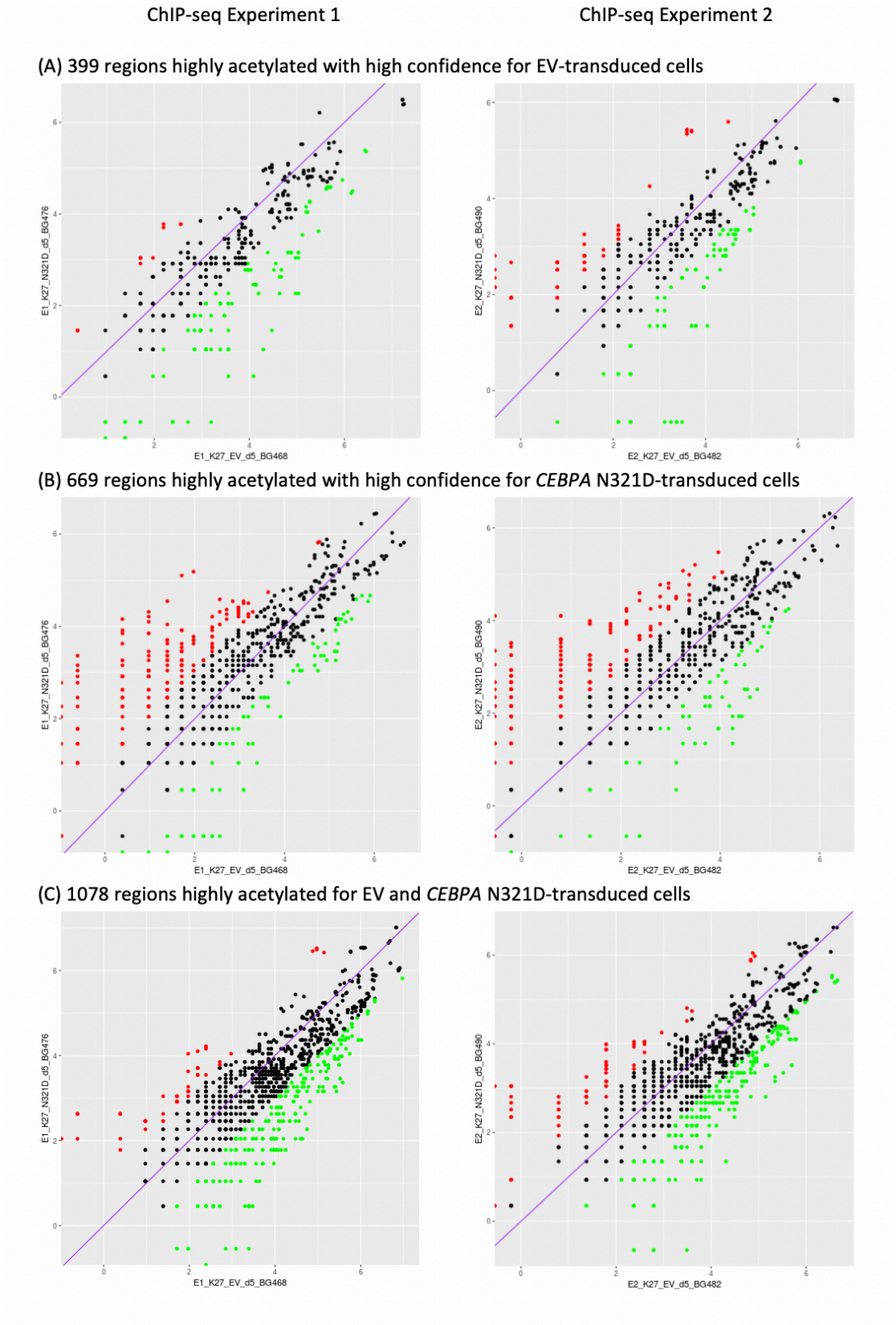


Figure 5.11 Regions enriched two-fold by H3K27Ac-binding

The following regions can then be mapped to genes: two-fold enriched in H3K27Ac in EV cells from the 399 regions in Figure 5.10, two-fold enriched in *CEBPA* N321D-transduced cells from the 669 regions, and two-fold enriched in EV cells from the 1078 regions. These are intersected between the two experiments to identify genes enriched for H3K27Ac with high confidence to give the following results (Table 5.2):

Table 5.2 High-confidence C/EBPα-bound genes associated with regions enriched two-fold for H3K27Ac

399 regions	21 genes in EV versus *CEBPA* N321D-transduced cells
669 regions	1 gene in *CEBPA* N321D versus EV-transduced cells
1078 regions	35 genes in EV versus *CEBPA* N321D-transduced cells

These genes are then intersected with genes derived from differential expression of EV versus *CEBPA*-transduced cells at Day 5 of differentiation on single cell RNA-seq, to identify genes where chromatin accessibility is associated with C/EBPα-binding and increased expression. This intersection gives only 1 high-confidence gene (*KLF6*) which is highly acetylated and bound by C/EBPα exclusively in EV-transduced cells and shows upregulation on differential expression of EV against *CEBPA* N321D-transduced cells at Day 5, and 10 genes (*CLEC4B1*, *PID1*, *PLEKHM3*, *SLC29A3*, *PLEKHF2*, *MARCH1*, *ATP2B1*, *RCBTB2*, *PDE7B*, *CTSC*) which are highly acetylated in both cell lines and upregulated on differential expression of EV against mutant cells at Day 5 of differentiation in Flt3L. There are no genes which are highly acetylated in *CEBPA* N321D-transduced cells and downregulated on differential expression.

5.8 Discussion

5.8.1 Choice of cell type and antibody for ChIP-seq

The experimental hypothesis for our ChIP-seq experiments is that the N321D mutation resides in the bZIP domain of C/EBPα and therefore can be predicted to alter coiled-coil heptad interactions, C/EBPα binding profile and function (Pogenberg *et al*, 2014). The choice of cell type and antibody was an important consideration in designing a ChIP-seq experiment to provide reliable insights into how the altered topology of C/EBPα N321D mediates transcriptional effects. Regarding the choice of *CEBPA* N321D and EV-transduced cells, retroviral vectors integrate preferentially into transcribed regions of the genome, and bearing in mind that TF ratios play an important role in phenotype, the addition of mutant alleles to

two wildtype alleles can reasonably be hypothesised to provide useful insights into the impact of mutant *CEBPA* on TF binding profiles. Theoretically a more useful comparator for the N321D model would be the *CEBPA* WT-transduced cell line, which would integrate an additional WT allele, but our early experiments demonstrated that the integration of *CEBPA* WT into the Hoxb8-FL cell line resulted in an altered initial phenotype in the presence of estradiol, meaning that the two cell lines started from a different phenotype. An additional experimental challenge regarding *CEBPA* WT-transduced cells was rapid differentiation and clonal selection so that it was difficult to obtain sufficient representative cells for ChIP-seq after 5 days of differentiation. One possible experiment would have been to compare binding profiles of *CEBPA* WT and EV-transduced cells at Day 0 before differentiation, but this would have failed to elucidate the leukaemic effects of the N321D mutation during differentiation. For these reasons, it was decided to perform ChIP-seq experiments to compare EV and *CEBPA* N321D-transduced cells at Days 0 and 5 of differentiation. It is also important to emphasise that this experimental model captures the effects of one specific C-terminal mutant isoform of C/EBPα. In fact, *CEBPA*-mutated leukaemia most commonly combines a C-terminal mutation with an N-terminal mutation, and recapitulation of this *in vivo* co-operative mutational profile would require a more complex cellular model with a second non-GFP reporter molecule and an alternative epitope tag, and then an analysis of which regulatory regions were bound by antibodies to both epitopes but not by C/EBPα antibody.

The other important decision in experimental design was antibody selection. The C/EBPα antibody was well-validated by our group and others, but it cannot discriminate between wildtype C/EBPα and the single altered amino acid in C/EBPα N321D. For this reason, it was decided to add a FLAG tag into the mutant plasmid construct and thereby to identify DNA regions bound specifically by the mutant TF rather than by wildtype. The FLAG tag is a hydrophilic polypeptide which can theoretically alter conformation and function of C/EBPα N321D. However, the tag is short (8 amino acids), and reassurance is provided by the fact that it was also included in the *CEBPA* WT construct, which showed a very different phenotype to the mutant, thus suggesting that the FLAG tag (shared between the two constructs) had no overriding dominant effect.

5.8.2 Qualitative changes are observed in binding events during differentiation of EV and *CEBPA* N321D-transduced cells

FLAG binding events increase in both experimental replicates during differentiation of *CEBPA* N321D-transduced cells, but interestingly there is no significant global increase in binding events for C/EBPα transcription factor during differentiation of either EV or *CEBPA* N321D-transduced cells (Figure 5.4). This suggests that closed chromatin opens up specifically to mutant C/EBPα N321D during differentiation, but that effects on wildtype C/EBPα binding are more subtle. It was therefore decided to subject C/EBPα binding profiles to more nuanced qualitative analysis by integrating ChIP-seq results with scRNA-seq expression profiles and secondly by looking at H3 Lysine 27 acetylation signatures to ask whether genes bound by C/EBPα in EV and *CEBPA* N321D-transduced cells can be distinguished by transcriptional activity.

Our previous gene expression analysis described genes which were dysregulated in EV and *CEBPA* N321D-transduced cell lines, however integrating these results with ChIP-seq allows us to determine genes which are direct targets of C/EBPα-mediated transcription. Two distinct approaches identify a common set of genes which can be considered with high confidence to be directly bound and transcribed by C/EBPα and which have altered gene expression in *CEBPA* N321D-transduced cells. These include genes which are associated with DC differentiation and function (*AXL*, *CD300a*, *ITGAM*, *TLR2*, *CTSB* and *NOTCH2*), with HSPC cell state (*TGFBI*, *MYB*), and with cell cycle control (*CXCR4*, *KLF4* and *E2F2*), suggesting overall that the *CEBPA* N321D mutation promotes a transcriptional signature which favours HSPC cell state and acts in opposition to DC differentiation. In addition, C/EBPα N321D TF binds directly to genes such as *SPI1*, *IRF8* and *RUNX1* which show no dysregulated gene expression in our experiments but which are known to encode important partner TFs that interact with wildtype C/EBPα to direct DC differentiation.

5.8.3 Intersection between C/EBPα-bound regions and H3K27Ac-bound regions identifies transcriptionally relevant genes

It is well-established that the majority of TF binding events do not have functional relevance in every tissue-type or cell-type, and therefore we employed H3K27Ac coverage to identify those regions where C/EBPα-binding was associated with transcriptionally active chromatin. One important caveat to interpreting the correlation between regions enriched for C/EBPα

and regions enriched for H3K27Ac is that the regulatory regions selected for by H3 Lysine 27 acetylation include enhancers which may be located many kilobases distant from the relevant transcribed gene. By contrast to enhancers, promoter regions are located around the transcriptional initiation site (Smale *et al*, 2003), bind RNA polymerase II, and can be selected by H3 Lysine 4 trimethylation (H3K4me3) (Sharifi-Zarchi *et al*, 2017). It would be interesting to conduct a future experiment with H3K4me3 ChIP antibody to compare promoter-driven binding profiles between EV and *CEBPA* N321D-transduced cells, particularly as these promoter regions would correlate more reliably with associated gene loci. Both H3K27Ac and H3K4me3 have been employed by others to analyse TF binding in dendritic cells (Wan *et al*, 2015). Nevertheless, the Göttgens group has experience of using the H3K27Ac ChIP antibody to identify binding sites which are transcriptionally active and therefore likely to have functional relevance.

When analysing regions in all cell lines which are bound by C/EBPα and which have increased coverage with H3K27Ac, we find these peaks to be enriched for intragenic and intergenic locations, suggesting that C/EBPα-binding regions with transcriptionally active chromatin are specific for enhancer rather than promoter regions. This corresponds with previous work by Heinz *et al* which showed that the majority of PU.1 and C/EBPα binding sites which localised to promoters demonstrated similar occupancy levels in B cell and macrophages, whereas differentially bound sites were located in distal enhancer sites (Heinz *et al*, 2010). However, intersection of C/EBPα-binding and H3K27 acetylation data with gene expression data from single cell RNA-seq identifies very few genes which demonstrate differential expression associated with increased acetylation in regions identified by C/EBPα peaks. This can possibly be explained by the three dimensional folding configuration of DNA, which means that an acetylated region or a C/EBPα peak may not correspond to the closest gene which is called by that peak. In this case, it is hypothesised that acetylation is happening at enhancers which are intergenic and so distal that peak-calling does not correctly identify genes which have increased transcriptional activity. This could be tested by chromosome conformation capture-on-chip methods such as 4C (Simonis *et al*, 2006) to identify the spatial distribution of chromatin in EV and *CEBPA* N321D-transduced cells. In addition, comparing acetylation profiles between the two conditions may not be informative where, for instance, H3K27 acetylation profiles are similar at a particular gene locus in both conditions, but C/EBPα binding at that locus happens preferentially in one condition, particularly when we remember

that C/EBPα is thought to be a pioneer factor which is capable of engaging nucleosomes and closed chromatin (Hasemann *et al*, 2014). Another challenge is that we cannot know whether C/EBPα-binding events in the *CEBPA* N321D-transduced cells are caused by wildtype or mutant C/EBPα. Lastly, it is reasonable to speculate that some of the transcriptional effects of the N321D mutation in the bZIP domain are associated not with DNA-binding profiles but rather with sequestering of cooperative transcription factors, and in this case transcriptional activity would not necessarily correlate with C/EBPα-binding events. It would certainly be interesting to employ co-crystallisation methods to study interactions at the LZ domain with TF partners, but such work is outside the scope of this thesis.

6. Restructuring, curating and expanding ChIP-seq datasets in CODEX database: New insights into cell-intrinsic and TF-mediated effects during human haematopoiesis

6.1 Introduction

Following the success of the human genome project, the scientific community has engaged in large-scale consortium projects such as ENCODE (ENCyclopedia Of DNA Elements) and ImmGen (Immunological Genome Project) to computationally reconstruct gene regulatory networks in different cell types. However, an alternative methodology has developed which builds up large databases by collating data from multiple small-scale projects and even individual experiments. Examples of this approach include Heat*seq which displays interactive correlation heatmaps from thousands of experiments using different cell types in human, mouse and drosophila (Devailly *et al*, 2016) and BloodSpot which includes gene expression profiles of normal and malignant haematopoiesis in murine and human cells (Machado *et al*, 2016). Similarly, SBR-Blood incorporates data from micro-array and NGS (RNA-seq, Methyl-Seq and ChIP-seq) experiments to describe cell-type specific analysis of regulatory mechanisms in mouse haematopoiesis (Lichtenberg *et al*, 2016). Specific databases have been constructed to analyse gene expression in stem cells, such as StemMapper (Pinto *et al*, 2018) which includes 166 transcriptomes generated on Affymetrix microarray technology and manually curated from NCBI's Gene Expression Omnibus (GEO). More recently, CancerSEA has collected RNA-seq data from 41,900 single cells for cancer-related scRNA-seq datasets on SRA, GEO and ArrayExpress (Yuan *et al*, 2019).

CODEX (Sanchez-Castillo *et al*, 2015) is a database constructed by the Göttgens group which collects raw data on transcription factor (TF)-binding and histone modification from multiple NGS experiments, processes them with a uniform standardised bioinformatics pipeline, and permits manual curation to ensure accuracy. Important features of the CODEX database are that it is user-friendly, includes a web-crawler which facilitates automated reports of relevant new samples from the GEO database, allows individuals to freely interrogate CODEX and compare data from their own experiments, and allows visualisation of data on the UCSC Genome Browser. CODEX is a database which has evolved from an original dataset of 53 ChIP-

seq samples (Hannah *et al*, 2011) and later more than 300 ChIP-seq studies from murine haematopoietic cells (Ruau *et al*, 2013). The latest iteration of CODEX includes data for mouse and human cells from over 1000 ChIP-seq, DNase-seq and RNA-seq samples, and is organized into two separate databases for blood cells (HAEMCODE) and embryonic stem cells (ESCODE) (Sanchez-Castillo *et al*, 2015).

The integration of TF-binding and histone modification data has provided novel insights into cooperative networks of haematopoietic gene regulation, for instance the finding that FLI1 and RUNX1 pair collaboratively in HSPCs (Beck *et al*, 2013) and that GATA2, GFI1B and FOS cooperatively induce haemogenic reprogramming in fibroblasts (Gomes *et al*, 2018). Our hypothesis is that an enhanced understanding of transcriptional networks regulating human haematopoiesis in general, and haematological malignancy in particular, can be achieved by curating the HAEMCODE database and inputting experimental data from more blood cell types.

6.2 Manual curation of human blood cells in HAEMCODE

The first stage of this process was manual curation of the existing database to ensure that all included cell types in HAEMCODE were blood cells, and that experimental data was annotated correctly.

Initial curation of 648 human datasets in the HAEMCODE database revealed a significant number (171) of non-haematopoietic datasets which were subsequently removed, including cervical and breast cancer as well as early embryonic tissues such as mesoderm and endoderm. The haematopoietic datasets included 18 RNAse and 13 DNAse experiments, which were excluded from further analysis. The remaining 446 experiments from HAEMCODE described TF-binding and histone modification. Curation involved detailed study of each cell type, subtype, tissue ontology from BRENDA database (BTO), experimental conditions, and additional details. For instance, experiments employing the K562 cell line were classified as erythrocytic leukaemia cells, but following curation CODEX users can now see additional detail that this cell line was originally cultured from a patient with CML in blast crisis and cells are positive for BCR:ABL mutation. In many instances, rigorous curation required literature review, for instance TF-binding experiments from GSE46044 used a Tet-inducible model of the

U937 cell line which was originally derived from a patient with histiocytic lymphoma and has been described by some authors as a "leukaemic monocyte lymphoma" cell line (Ishikawa *et al*, 2012). However, literature review revealed that it is employed as an immortalised monocyte progenitor cell which can differentiate into functional macrophages (Sproston *et al*, 2018), and therefore was recategorised as a monocyte in the HAEMCODE database. Literature review proved an effective method for curation, although it was difficult to categorise a ChIP-seq experiment in one instance because it used the Ficoll-Paque method to isolate peripheral blood mononuclear cells and this method cannot distinguish between mononuclear cells which include T-cells, B-cells, NK cells and monocytes (Wang *et al*, 2013).

By employing the CODEX web crawler and by manually searching GEO for new human haematopoietic cell lines, with a specific emphasis on haematological malignancy, we then identified additional 197 datasets. The previous published data (A) and the new collated data (B) is summarised below in Figure 6.1.

Since less than 5% of the genome encodes proteins, and many mutations and single nucleotide polymorphisms (SNPs) map to the non-coding portion of the genome, there is an urgent need to annotate non-coding DNA. A key metric that can be employed to characterise the utility of a data resource of ChIP-Seq experiments is the fraction of the genome for which it provides annotation, calculated by identifying total bases with non-zero coverage for all TFs in all human samples in the database. In terms of TF binding data, annotation of the human genome has increased from 9.94% since CODEX was previously published (Sanchez-Castillo *et al*, 2015) to 14.42% after our curation and expansion of the human HAEMCODE database. Of note, this is 3-fold higher than the portion of the genome that is coding, and therefore of potentially broad interest to the wider haematology research community. The utility of identifying regions of TF binding is that it provides evidence of regulatory circuitry at the locus in question, and additionally this information can suggest transcriptional relevance if mutations or SNPs are identified in non-coding DNA by genome wide association studies (GWAS).

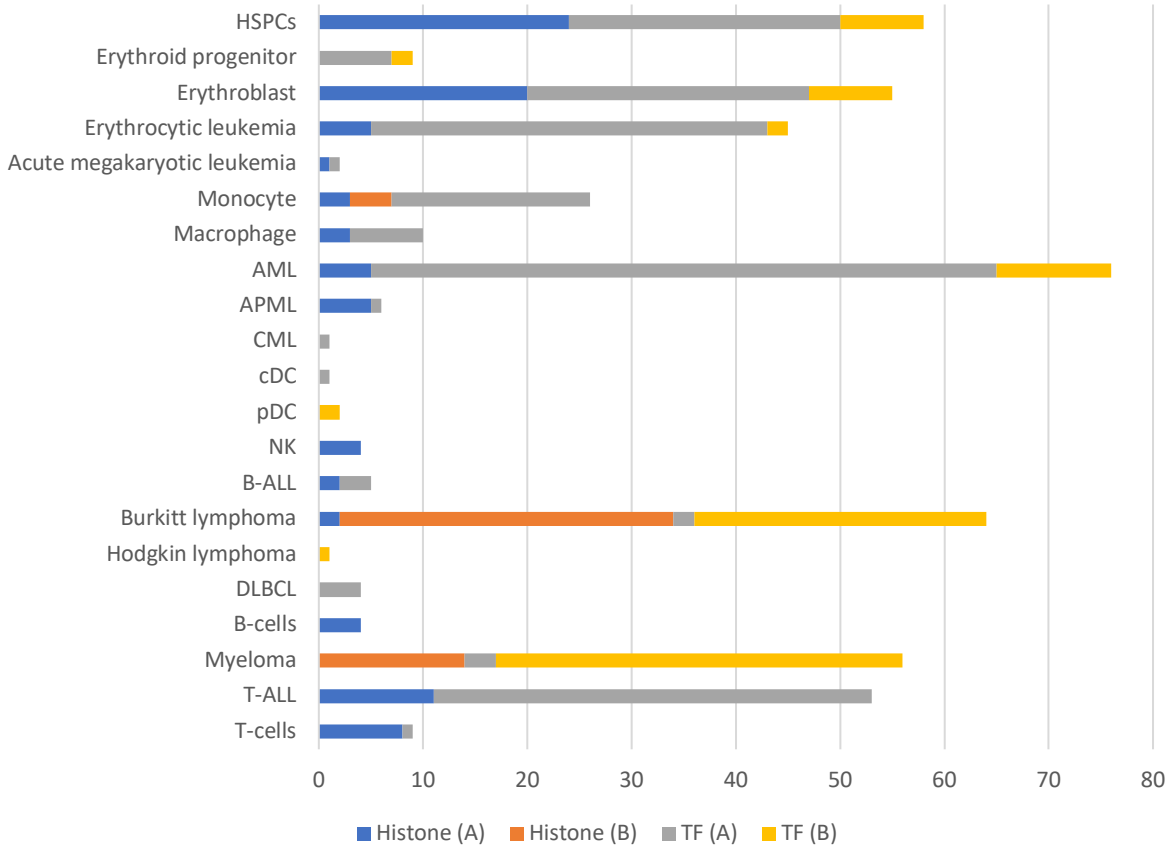


Figure 6.1 Summary of numbers of ChIP-seq experiments in (A) previous iteration of CODEX and (B) additional data incorporated during my review

6.3 Constructing global heat maps for human blood cells reveals pairwise correlations between cell types in TF-binding data but not in histone modification data

Heat maps were generated by R Hannah from the Göttgens lab. Briefly, she combined peaks from all TF samples and generated heat maps where hierarchical clustering was based on using a DICE coefficient matrix to establish pairwise correlation between binding patterns across regions. This was a binary process based on presence / absence of each peak, and linking these peaks to specific genomic loci using Model-based Analysis of ChIP-Seq (MACS) (Zhang *et al*, 2008).

The first heat map was generated using TF data from all cell types:

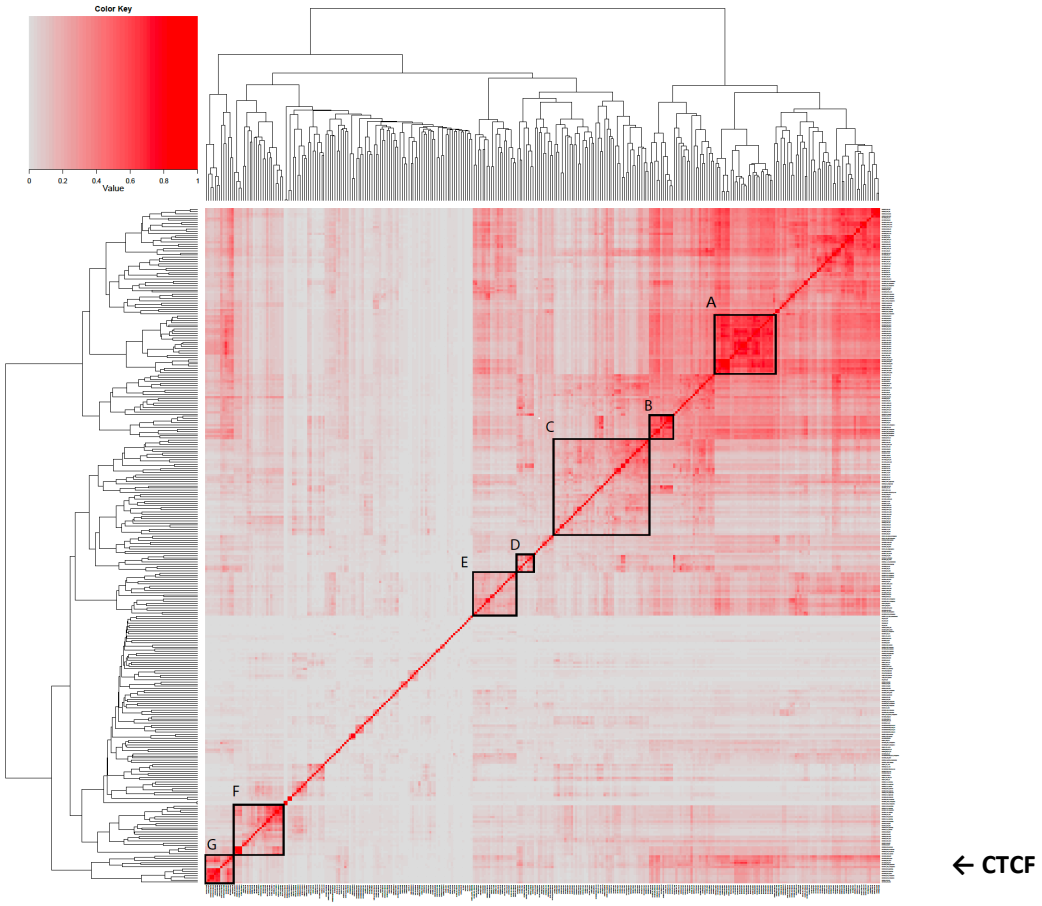


Figure 6.2 Heat map showing pairwise correlations between ChIP-seq profiles for transcription factors in all human blood cells on CODEX (A) is myeloma cells, (B) is T-cell lymphoma cell lines, (C) is AML and HPC cells, (D) includes macrophages and dendritic cells, (E) is diverse in cell type and TF identity, (F) is erythroblasts, (G) is diverse cell types with over-representation of **CTCF**

This first heat map permitted several interesting observations. Firstly, CODEX creates distinct TF clusters on heat maps, for instance clusters A, C and F in Figure 6.2 denote myeloma cells, AML and haematopoietic precursors cells, and erythroblasts respectively. Secondly, the unified standardised CODEX pipeline is capable of clustering TF binding data for similar cell types from distinct experimental sources, for instance Figure 6.2B clusters together Jurkat cells and CUTLL1 cells which are both distinct T-cell lymphoma cell lines. Thirdly, there are clusters where diverse cell lines cluster together with a shared TF, for instance cluster G is a cluster of myeloma, erythroid progenitor, Burkitt lymphoma, haematopoietic precursor cell, and K562 cell lines based on binding patterns from the shared transcription factor CTCF. CTCF is a highly conserved 11 zinc finger TF which binds to about 60,000 sites in the human genome (Chen *et al*, 2012) and there is evidence that it plays important roles in HSCs (Kim *et al*, 2017), myeloid cells (Ouboussad *et al*, 2013) and lymphoid cells (Heath *et al*, 2008); cluster G suggests

that the global binding profile of this TF is able to create pairwise correlations across various cell lines. This suggests that CTCF binding, at least in part, is similar across different haematopoietic lineages, perhaps in agreement with its role in establishing/maintaining coarse-grained chromatin domains.

By contrast, cluster E in Figure 6.2 denotes pairwise correlations between diverse cell lines with different TFs that nonetheless show pairwise correlation in their chromatin binding profiles. Lastly, cluster D shows a pairwise correlation between the TF binding patterns of macrophages, plasmacytoid dendritic (pDC) and conventional dendritic (cDC) cells, with the pDC ChIP-seq profile having a closer pairwise correlation with lymphoid experiments and the cDC profile being more closely correlated with myeloid experiments. Overall, these results show that CODEX's standardised bioinformatics pipeline is capable of bringing together raw data from multiple ChIP-seq experiments carried out under different conditions by different centres, and that pairwise correlations can be found based both on cell type and TF.

The Hardison group's S3norm method (Xiang *et al*, 2018) was then employed to infer a heat map of pairwise correlations between histone modification profiles in HAEMCODE (Figure 6.3). By contrast with TF binding profiles, which are commonly characterised in binary terms as peak present or peak absent, enrichment for a given histone modification is generally thought of as a more quantitative process, and therefore it is more challenging to deal with sequencing depth (SD) and signal-to-noise ratio (SNR) when modelling chromatin accessibility landscapes. S3norm employs nonlinear transformation to simultaneously normalise both SDs and SNRs between data sets (Xiang *et al*, 2018), and therefore provides superior performance for H3K4me3 ChIP-seq profiles in comparative studies of normalisation, particularly with reference to SNR.

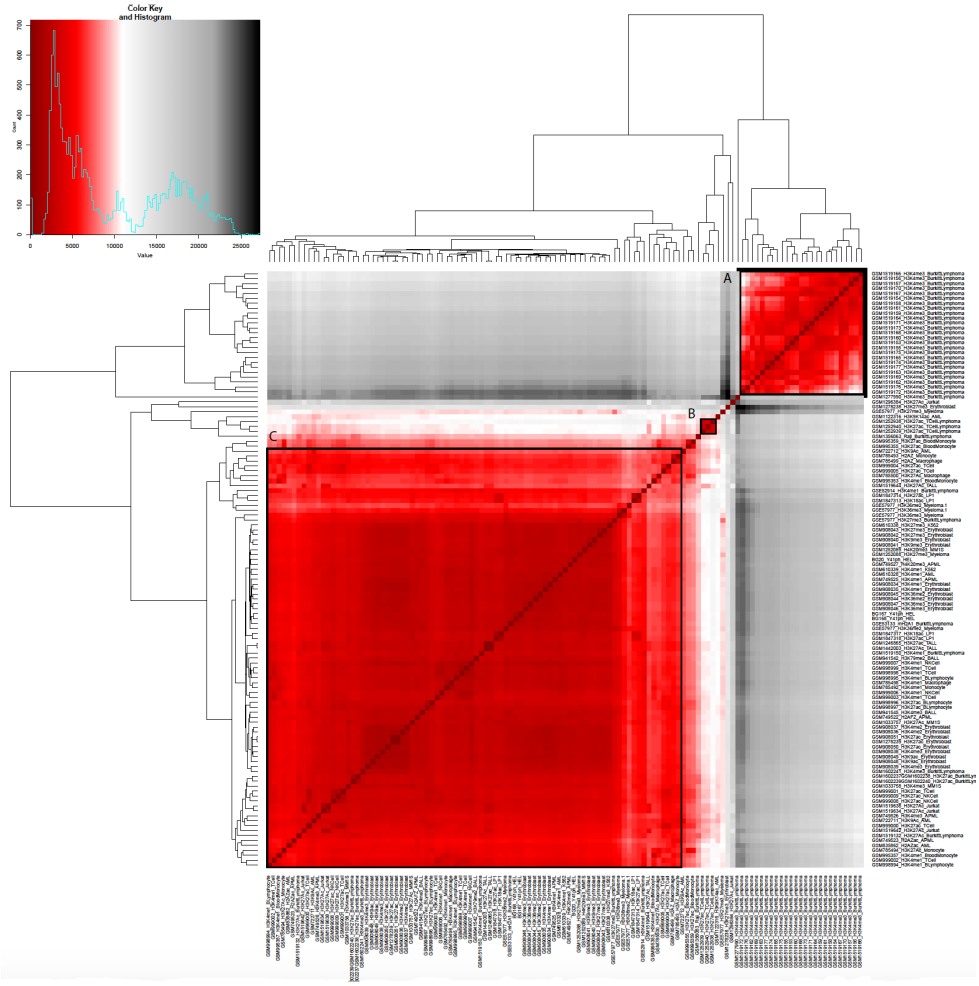


Figure 6.3 Heat map showing pairwise correlations between ChIP-seq profiles for histone modifications in all human blood cells on CODEX (A) is Burkitt Lymphoma cells, (B) is T-cell lymphoma cell lines, (C) is diverse cell lines with different histone markers

Initial correlation analysis of histone modification experiments from HAEMCODE is shown in Figure 6.3, and gave a very distinct cluster of Burkitt Lymphoma cells bound by H3K4me3. For this reason, it was decided to remove the Burkitt cell line experiment from the histone heat map and rerun pairwise correlations using the S3norm method to uncover any masked correlations between histone modifications for other cell types (Figure 6.4):

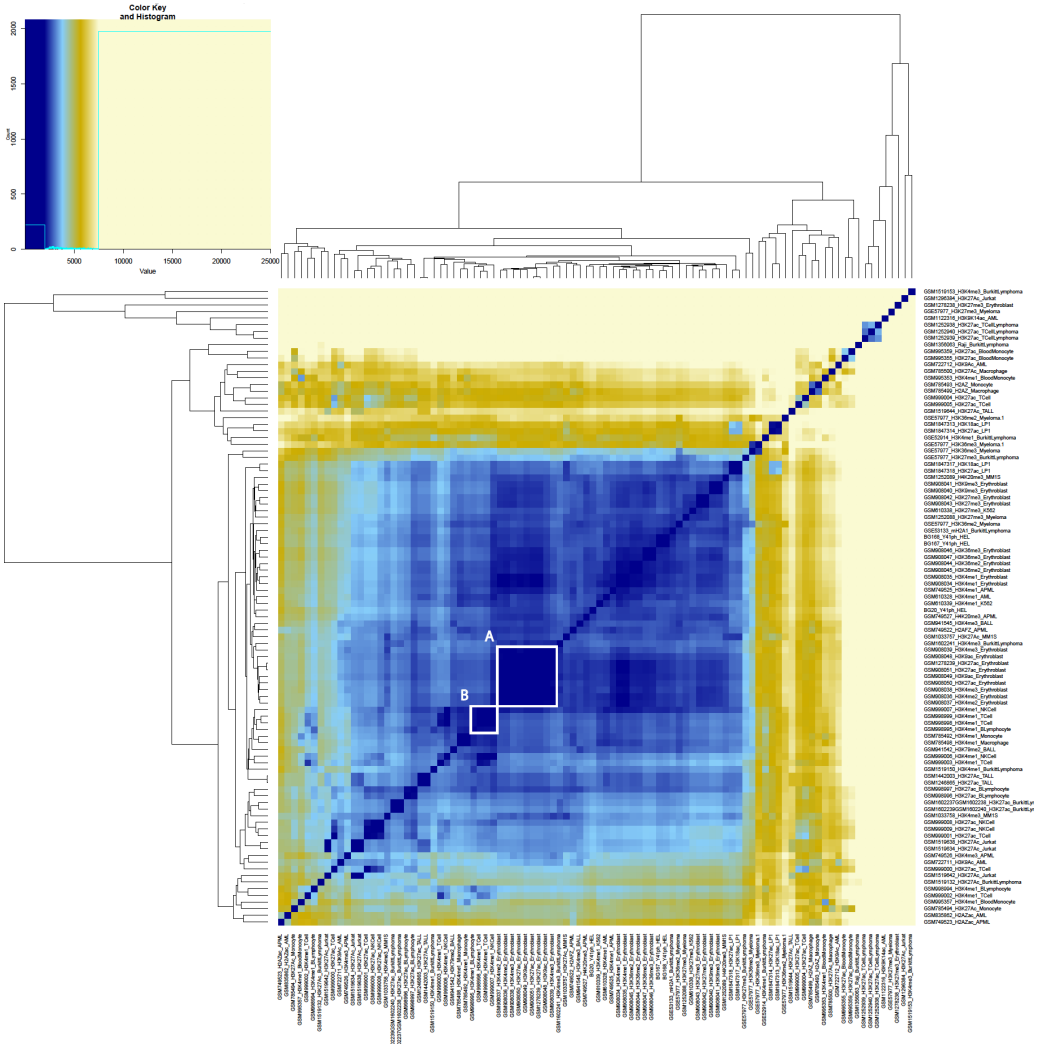


Figure 6.4 Heat map showing pairwise correlations between ChIP-seq profiles for histone modifications in all human blood cells on CODEX after excluding experiments from cluster A in Figure 6.3 (A) is erythroblasts, (B) is NK, T-cell, B-cell and monocyte cell lines

By contrast with TF-binding experiments, histone modification experiments do not cluster sharply by pairwise correlations. Of note, the Burkitt Lymphoma cluster A in Figure 6.3 is derived from one experimental protocol and employs only one histone marker, whereas other experiments with Burkitt Lymphoma cells do not correlate on the heat map. This may be because of the aforementioned challenges related to peak-calling with chromatin accessibility data.

6.4 Constructing heat maps for normal and malignant human blood cells reveals distinct clusters in normal human blood cells but loss of pairwise correlation in malignant cells

We then employed the DICE coefficient matrix to look at subsets of the TF data, initially by looking separately at normal (Figure 6.5) and malignant (Figure 6.6) human blood cells:

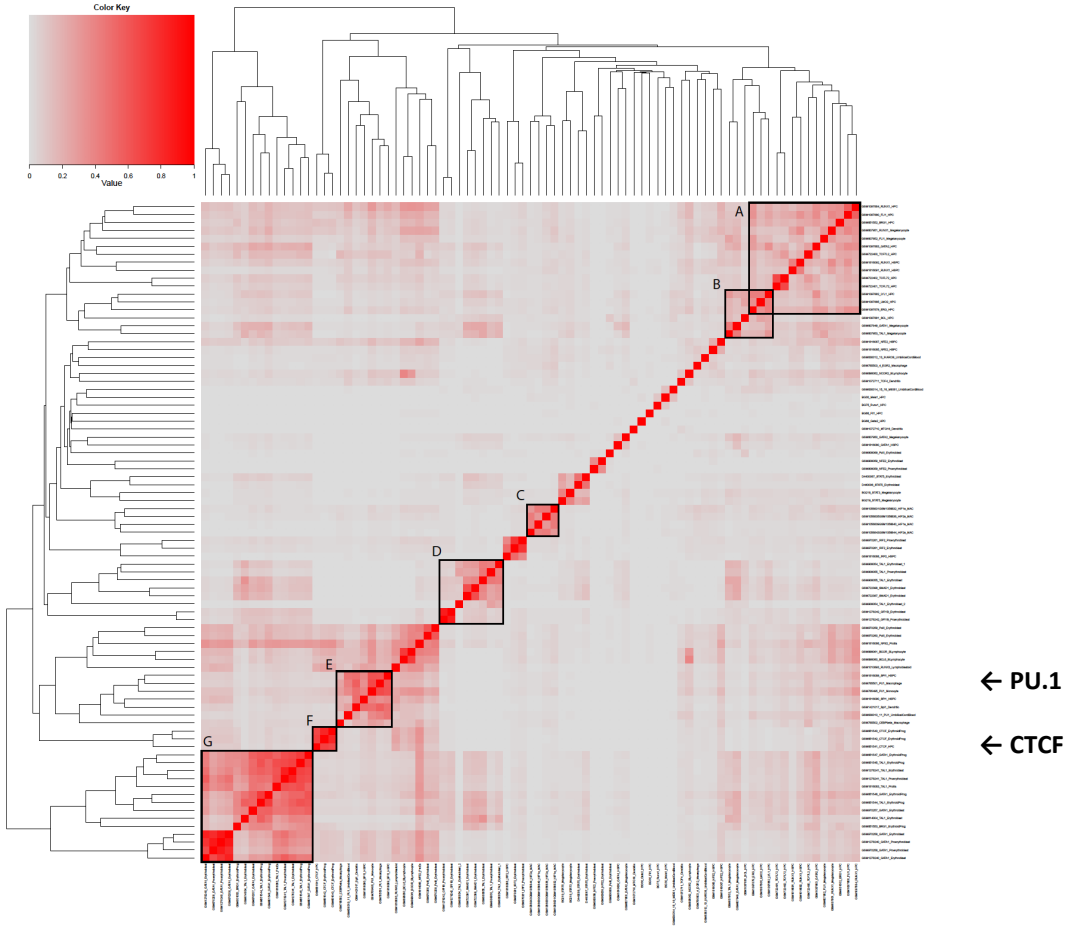


Figure 6.5 Heat map showing pairwise correlations between ChIP-seq profiles for transcription factors in non-malignant human blood cells on CODEX (A) and (B) are mostly HPCs and HSPCs with some megakaryocyte cells, (C) is macrophages, (D) is erythroblasts, (E) is macrophages, monocytes, and dendritic cells with **PU.1**, (F) is HPCs and erythroid progenitors with **CTCF**, (G) is pro-erythroblasts, erythroblasts and erythroid progenitors

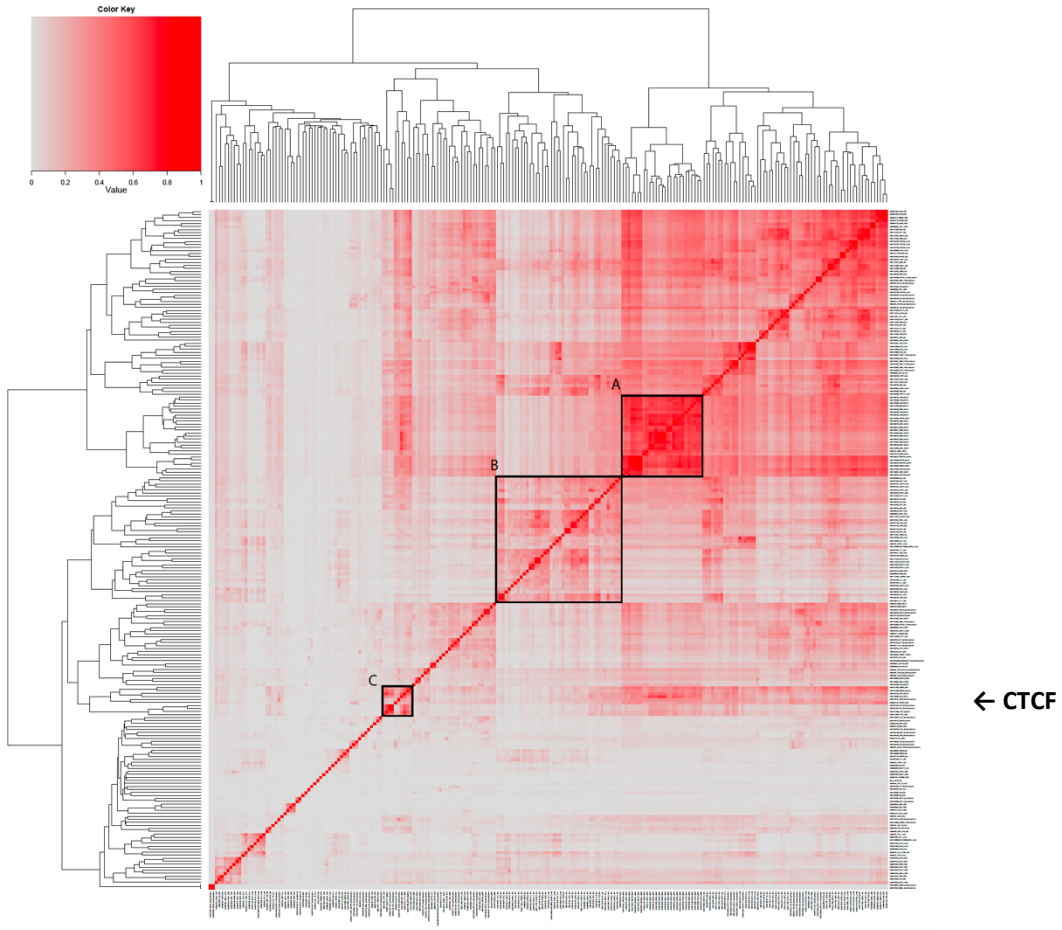


Figure 6.6 Heat map showing pairwise correlations between ChIP-seq profiles for transcription factors in malignant human blood cells on CODEX (A) is MM.1S multiple myeloma cell line, (B) is mostly AML but a few T-ALL and Jurkat cells, (C) includes Burkitt Lymphoma, myeloma and K561 cell lines with over-representation by **CTCF**

Non-malignant cells in Figure 6.5 pair into discrete clusters which are functionally and biologically distinct, ranging from HPCs to erythroid cells. For instance, in cluster E several cells with immune function cluster together. These pairwise correlations are much more difficult to elucidate on TF peak profiles in malignant cells (Figure 6.6), and we can hypothesise that this may be because the normal function and transcriptional wiring of these cells is disrupted. The only cells which form a distinct and unique cluster are myeloma cells, which despite somatic hypermutation and phenotypic diversity are positioned in a very distinct plasma cell compartment (Hansmann *et al*, 2017) within the haematopoietic tree. Intriguingly, some AML cells cluster with T-ALL cells, which may reflect a common transcriptional program associated

with aggressive proliferation. In both normal and malignant cells, CTCF seems to exert a particularly powerful effect on TF binding profiles across cell lines.

6.5 Constructing heat maps for progenitor, myeloid and lymphoid compartments suggests altered TF binding profiles for different cell compartments, and an association between AML and pDCs

Given their phenotypic similarity and early developmental stage, progenitor cells demonstrated no obvious clusters based on HSPC, HPC or CD133^{+} umbilical cord blood cell types, although increased pairwise correlation can be seen in cells where DNA is bound by ERG, LYL1, and LMO2 (cluster A) or PU.1 (cluster B). One can speculate that TF effects are more prominent than cell-intrinsic effects at an early stage of differentiation before lineage commitment and definitive transcriptional programs have initiated.

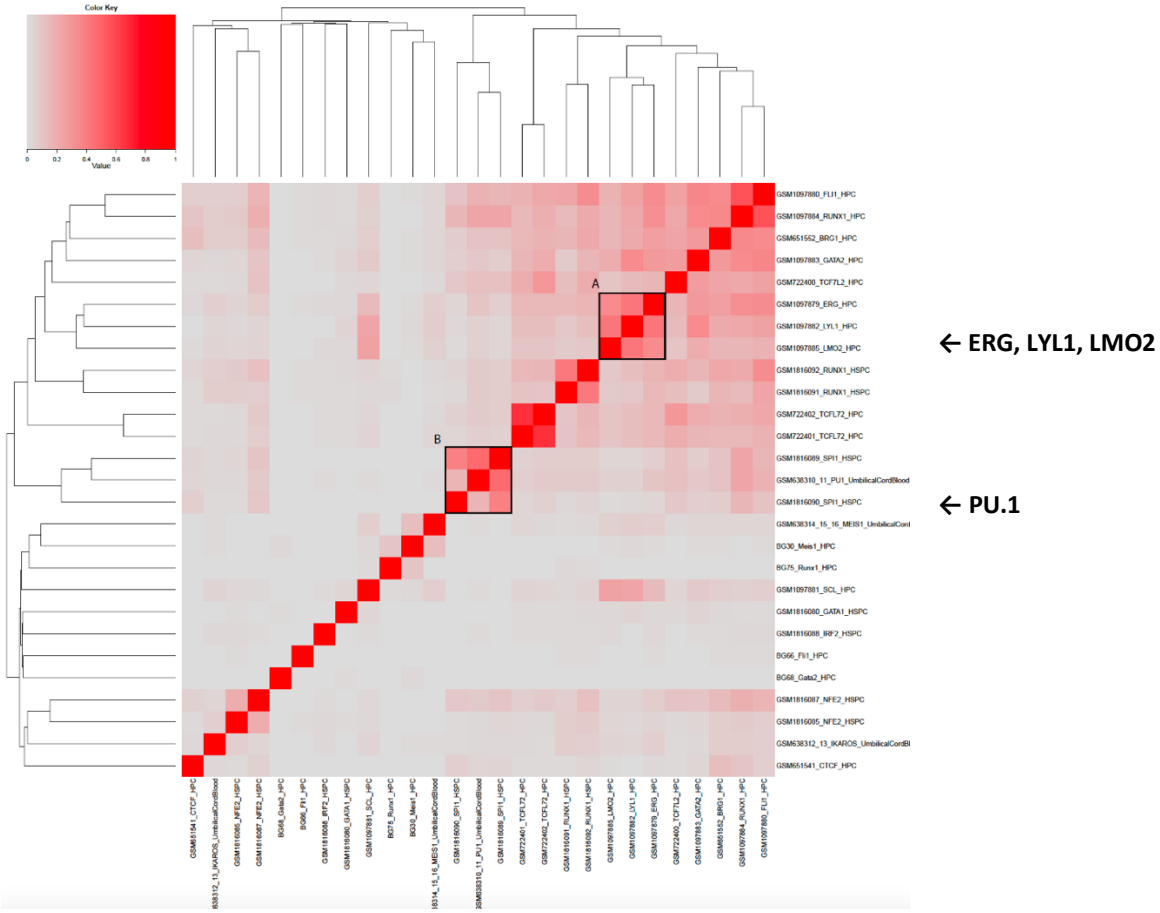


Figure 6.7 Heat map showing pairwise correlations between ChIP-seq profiles for transcription factors in human haematopoietic progenitor cells on CODEX (A) is HPC cells where DNA is bound by **ERG**, **LYL1** and **LMO2**, (B) is HSPC and CD133^{+} umbilical cord blood cells with **PU.1**

By contrast, ChIP-seq TF binding profiles for myeloid (Figure 6.8) and lymphoid (Figure 6.9) cells demonstrate pairwise correlation based on cell lines, with particularly prominent clusters for AML, Burkitt Lymphoma, T-cell lymphoma and myeloma, although a region within cluster B in Figure 6.9 shows intense pairwise correlation for cells where DNA is bound by MYB, and cluster C shows that CTCF continues to exert a powerful effect on overall binding profile across cell lines.

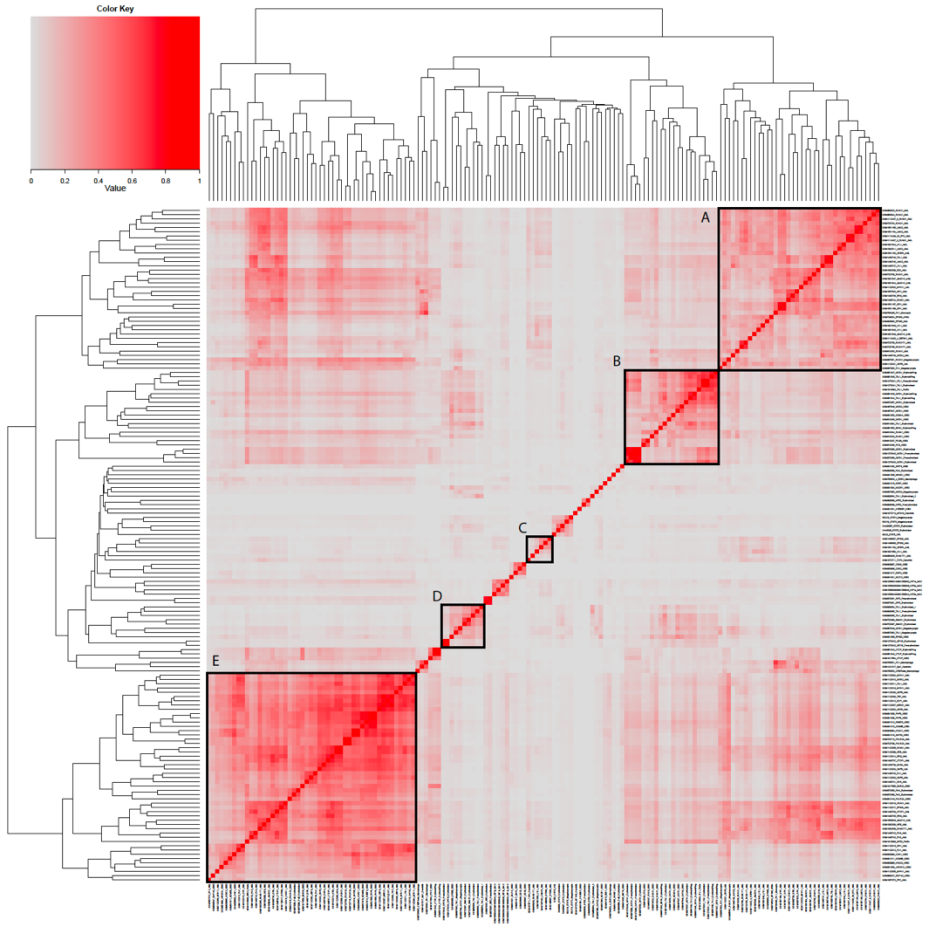


Figure 6.8 Heat map showing pairwise correlations between ChIP-seq profiles for transcription factors in human myeloid cells on CODEX (A) is AML cells, (B) is pro-erythroblasts, erythroblasts, erythroid progenitors and K562 cells, (C) is diverse cells including AML cells with **C/EBPα** and dendritic cells with **TCF4**, (D) is erythroid cells and megakaryocytes, (E) is mostly AML and K562 cells but also includes two erythroblast cells

Intriguingly, cluster C in Figure 6.8 shows pairwise correlation between AML cells where DNA is bound by C/EBPα and plasmacytoid dendritic cells where DNA is bound by TCF4. Transcription Factor 4 (TCF4) is a master regulator in pDC differentiation and is a highly

sensitive marker for blastic plasmacytoid dendritic cell neoplasm (BPDCN) (Khoury, 2018). Unfortunately, my own ChIP-seq data from the *CEBPA* N321D murine cell line cannot directly be compared to human HAEMCODE data, but it is nonetheless interesting to observe this correlation.

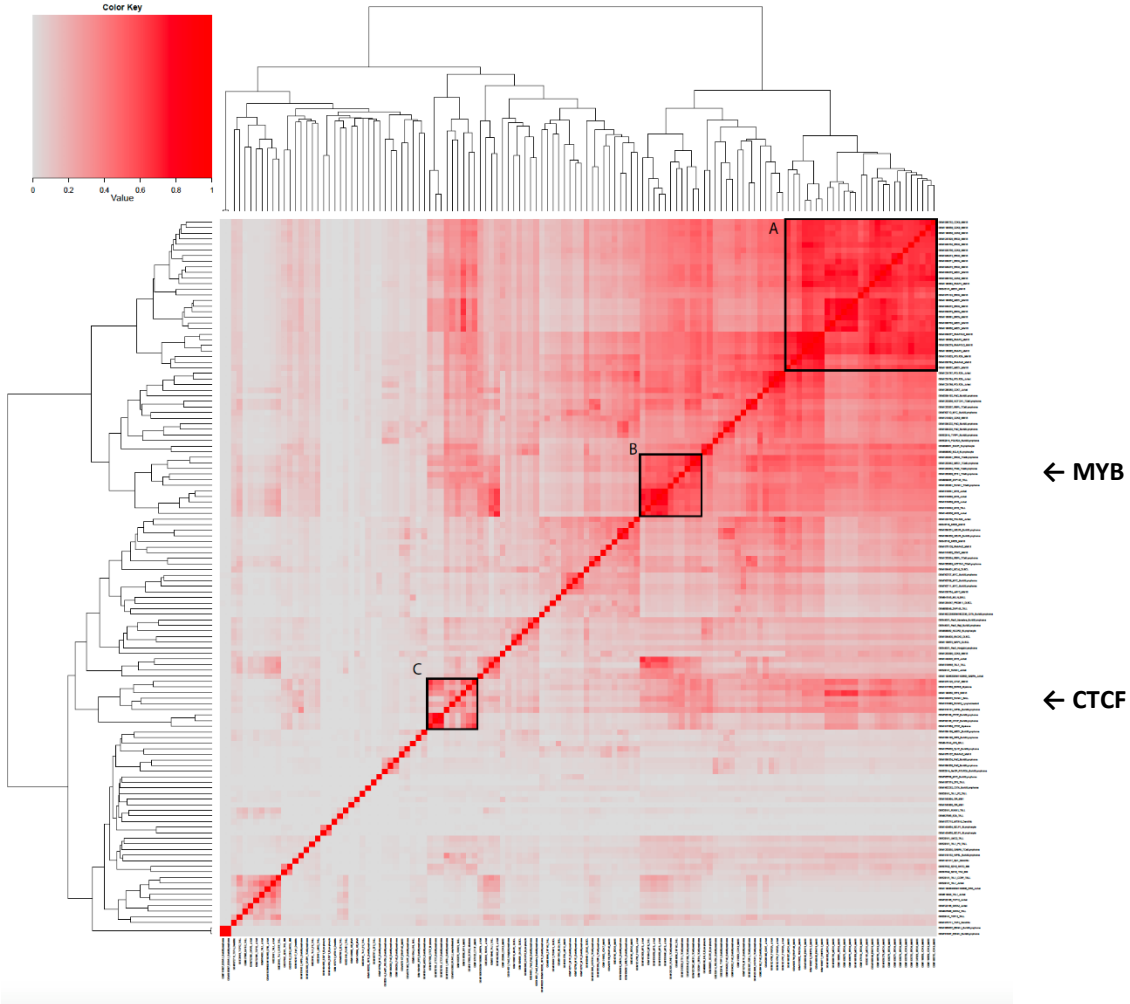


Figure 6.9 Heat map showing pairwise correlations between ChIP-seq profiles for transcription factors in human lymphoid cells on CODEX (A) is MM.1S cells, (B) T-cell lymphoma, T-ALL and Jurkat cells with **MYB**, (C) is Burkitt Lymphoma and myeloma cell lines with **CTCF**

6.6 Constructing heat maps for myeloid and lymphoid subsets within normal and malignant compartments

These subsets offer more granular data on the interaction between cell-intrinsic effects and TF-specific effects in ChIP-seq binding profiles on CODEX. For instance, in normal myeloid cells (Figure 6.10), we can observe cell-intrinsic effects for macrophages in cluster B, but we can also see TF-specific effects across cell lines with STAT5 and PU.1 in clusters A and C respectively. Clusters D and E exemplify this interaction, with cluster E forming a GATA1-specific subset within the erythroid progenitor cells in cluster D. The PU.1:GATA1 paradigm has been well-described (Graf *et al*, 2009), this paradigm has been recently challenged by employing novel reporter mouse lines and live imaging for continuous single cell long-term quantification of PU.1 and GATA1 transcription (Hoppe *et al*, 2016), but GATA1 remains an essential driver of erythroid differentiation which forms a distinct cluster on CODEX.

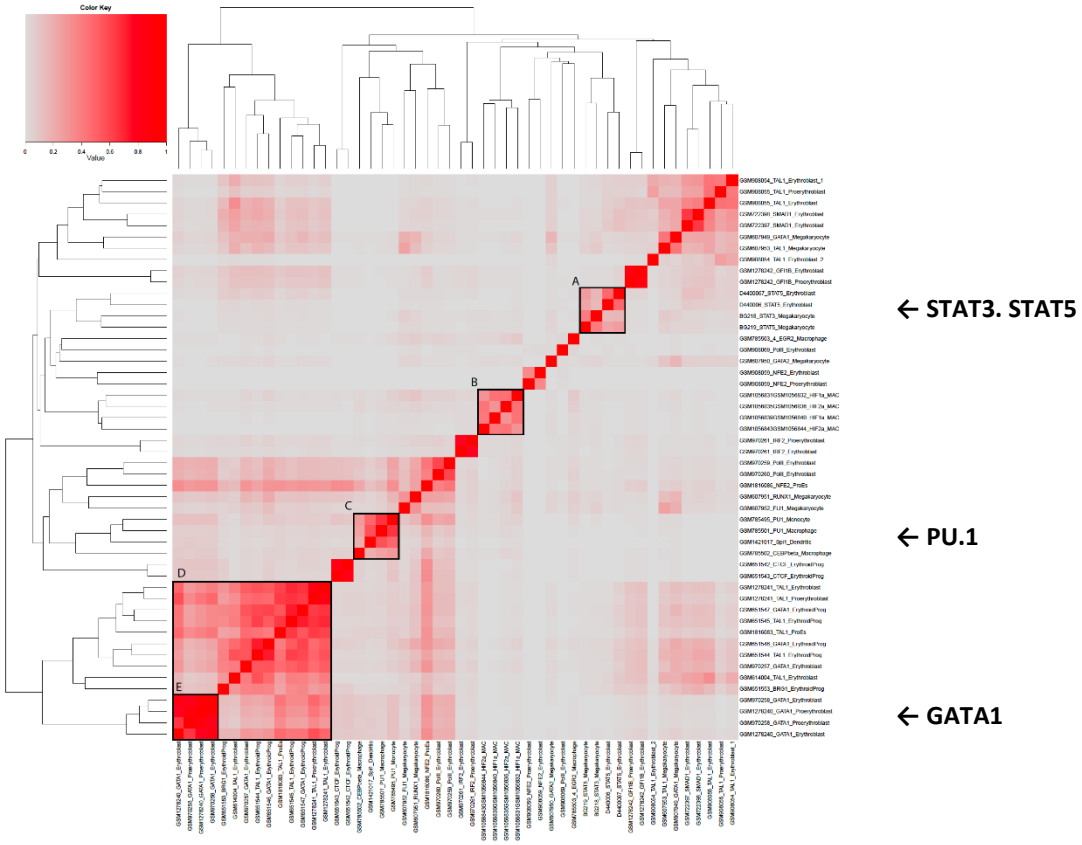


Figure 6.10 Heat map showing pairwise correlations between ChIP-seq profiles for transcription factors in normal myeloid cells on CODEX (A) is erythroblasts and megakaryocytes with **STAT5** or **STAT3**, (B) is macrophages, (C) is macrophages and cDCs with **PU.1**, (D) is pro-erythroblast, erythroblast and erythroid progenitors, (E) is a subset with **GATA1** only

Figure 6.11 describes pairwise correlations between ChIP-seq binding profiles for normal lymphoid cells. The distinct clusters A and C suggest that B lymphocytes and plasmacytoid dendritic cells have distinct binding profiles, however it is important to state that the small number of samples preclude any definitive conclusions.

It is worth stating here that the developmental biology of cDCs and pDCs remains a controversial question, with evidence that these cells can develop through both myeloid and lymphoid differentiation pathways. Our decision to place our cDC ChIP-seq profile in the normal myeloid compartment (Figure 6.10) and our pDC ChIP-seq profile in the normal lymphoid compartment (Figure 6.11) was based on our previous observations under the heat map in Figure 6.2 which examined TF binding profiles for all cells together. This is an example of how CODEX can be employed to define cryptic cell lines by pairwise correlation.

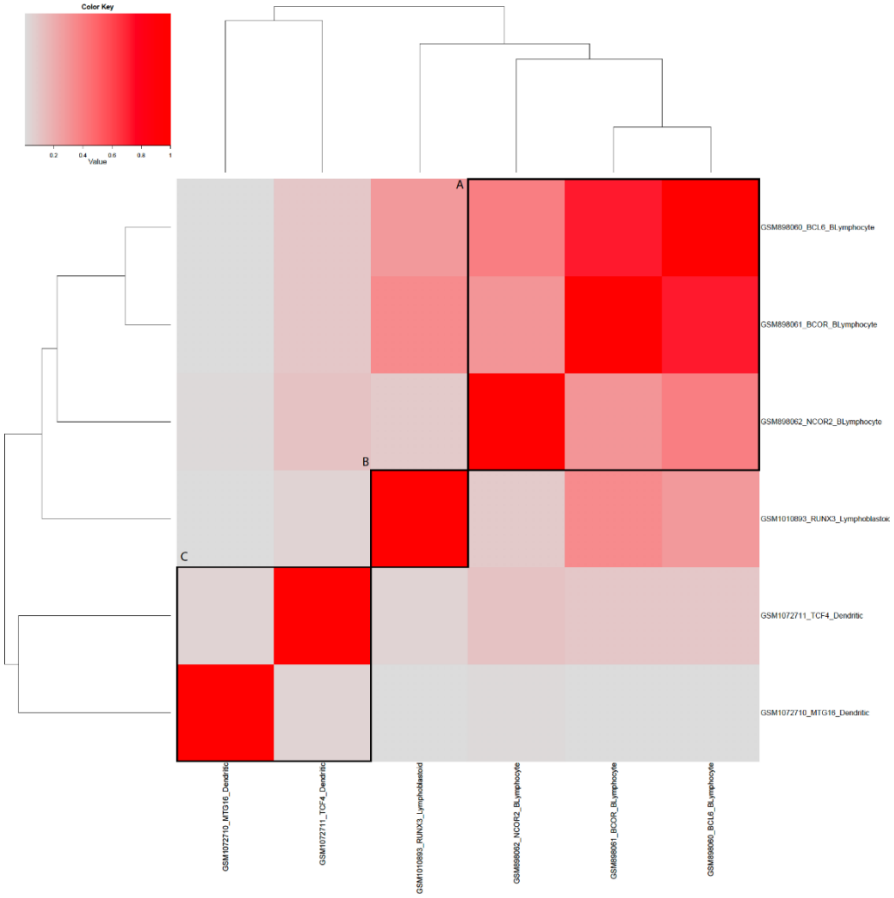


Figure 6.11 Heat map showing pairwise correlations between ChIP-seq profiles for transcription factors in normal lymphoid cells on CODEX (A) is B lymphocytes, (B) is lymphoblastoid cells, (C) is plasmacytoid dendritic cell

Figure 6.12 shows pairwise correlations between malignant myeloid cells. It is difficult to draw conclusions because this subset includes only AML and K562 cell lines, however we can observe that the most distinct clusters A and C are formed by AML cells, and that cluster B is formed by K562 cells which are driven by master regulator TFs such as GATA1, RUNX1 and FOSB.

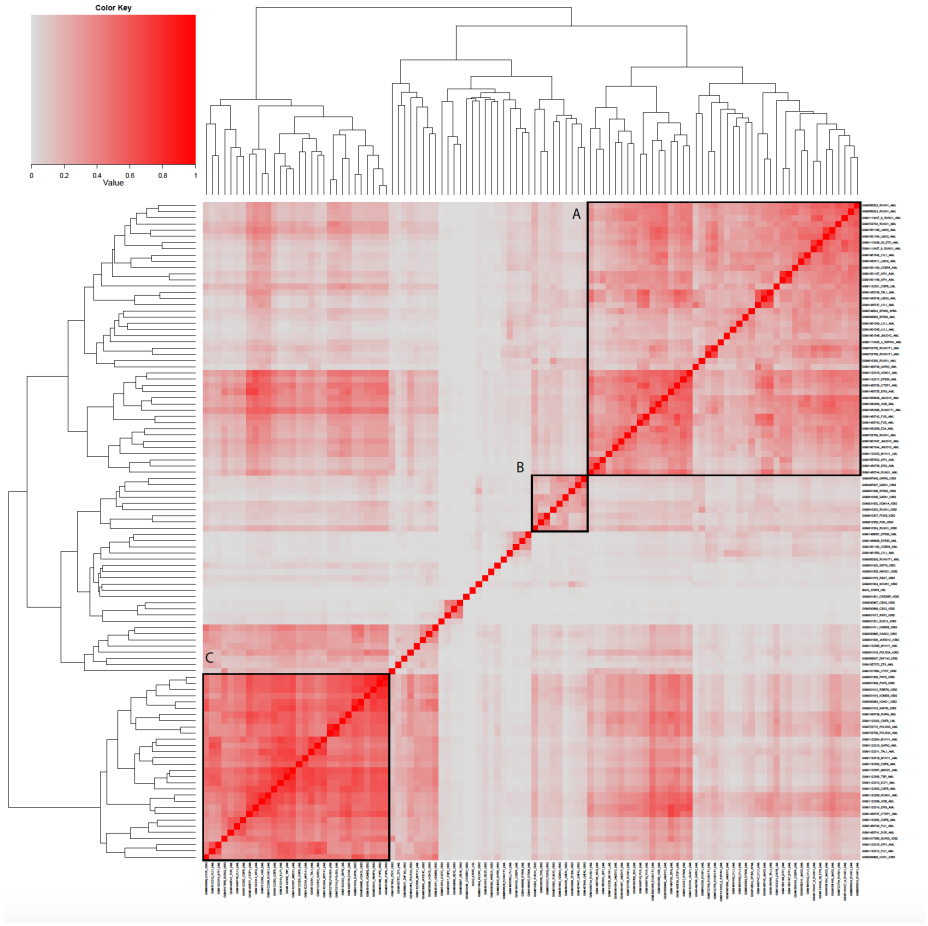


Figure 6.12 Heat map showing pairwise correlations between ChIP-seq profiles for transcription factors in malignant myeloid cells on CODEX (A) is AML, (B) is K562 cells, (C) is mostly AML cells with some K562 cells

Figure 6.13 confirms previous observations that distinct clusters are formed by binding profiles for myeloma cells (cluster A), and T-ALL and Jurkat cells (clusters B and C), and that CTCF-binding characterises pairwise correlation across Burkitt Lymphoma and myeloma cell lines (cluster D).

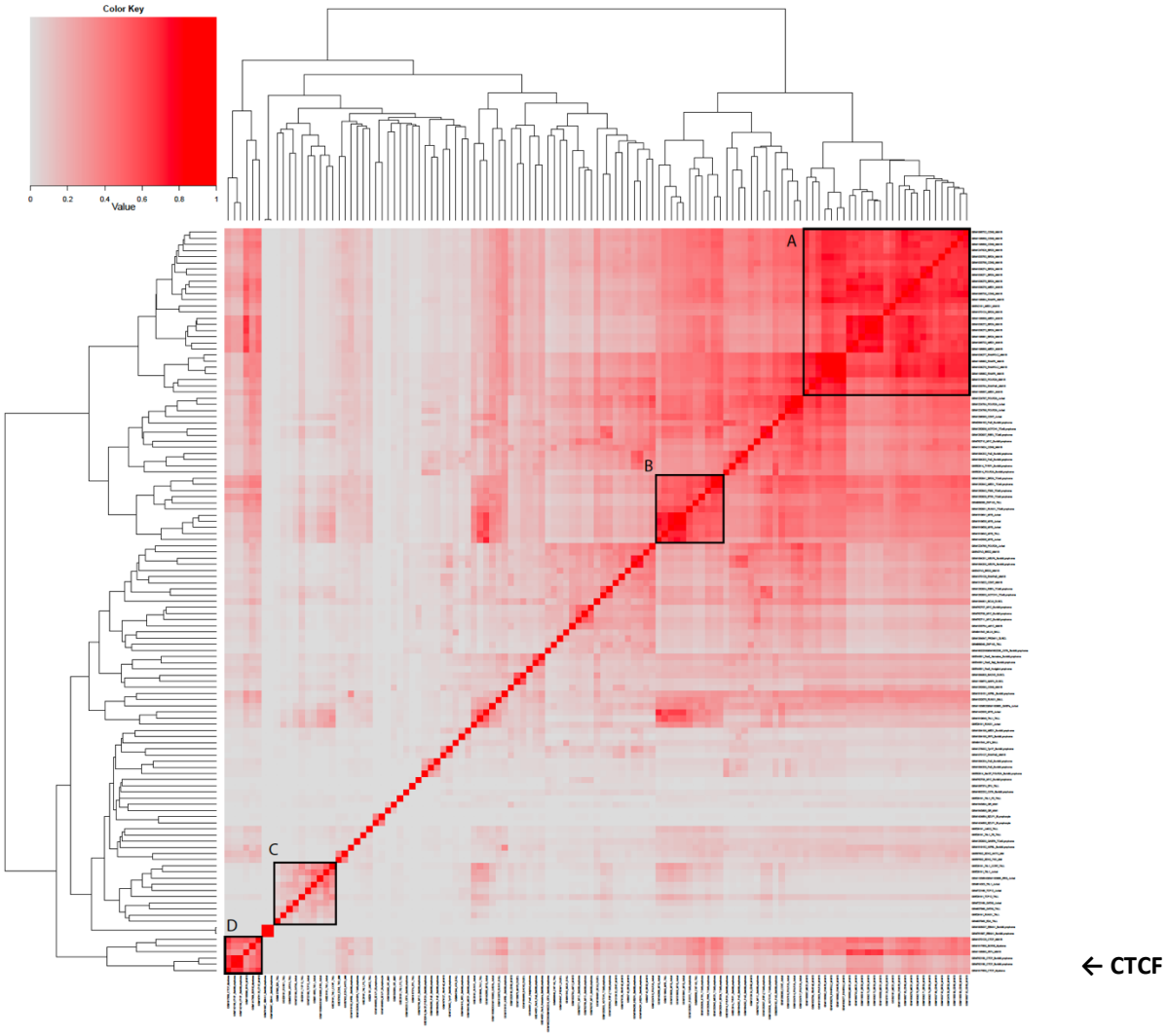


Figure 6.13 Heat map showing pairwise correlations between ChIP-seq profiles for transcription factors in malignant lymphoid cells on CODEX (A) is MM.1S cell line, (B) is T-ALL and Jurkat cells, (C) is T-ALL and Jurkat cells, (D) is Burkitt Lymphoma and myeloma cell lines with **CTCF**

6.7 Discussion

Several observations can be made through the HAEMCODE database of human blood cells. Firstly, TF DNA binding profiles form more distinct clusters on pairwise correlation of peaks than are formed by histone binding profiles, even when the S3norm protocol is employed to normalise data and account for SNR. Secondly, TF identity seems to exert more influence on pairwise correlations for TF binding profiles in progenitors, but cell-intrinsic correlations are more evident in lineage committed cells. Thirdly, TFs do nonetheless continue to influence binding profiles within clusters of similar cells, for instance we observe that GATA1 forms a distinct sub-cluster within the erythroid progenitor cluster. Fourthly, pairwise correlation of

TF DNA-binding profiles can reveal surprising correlations which merit further exploration, such as the correlation between AML cell lines and T-ALL cell lines. Fifthly, clusters are more difficult to establish in malignant cells than in normal cells, and this may be related to disruption of defined biological processes in normal cells. Lastly, CTCF has a singular capacity to maintain a consistent binding profile between different cell types in both normal and malignant cells, which may be related to its effects on chromatin looping and architecture (Hanssen *et al*, 2017).

CODEX is a powerful tool to probe DNA-binding profile data. For instance, pairwise correlations of human blood cells suggests that our pDC binding profile is closely related to lymphoid cell lines and our cDC binding profile is related to myeloid cell lines.

In addition, CODEX is able to harness experimental data from multiple sources and infer novel biological relationships. For instance, one of our findings is that CTCF has a very similar DNA binding profile in cell types as diverse as myeloma and Burkitt Lymphoma. It is known that CTCF has several binding sites within the *NANOG* locus, and therefore UCSC genome browser can be employed to visualise CODEX data for CTCF binding at this locus for myeloma and Burkitt Lymphoma cells:

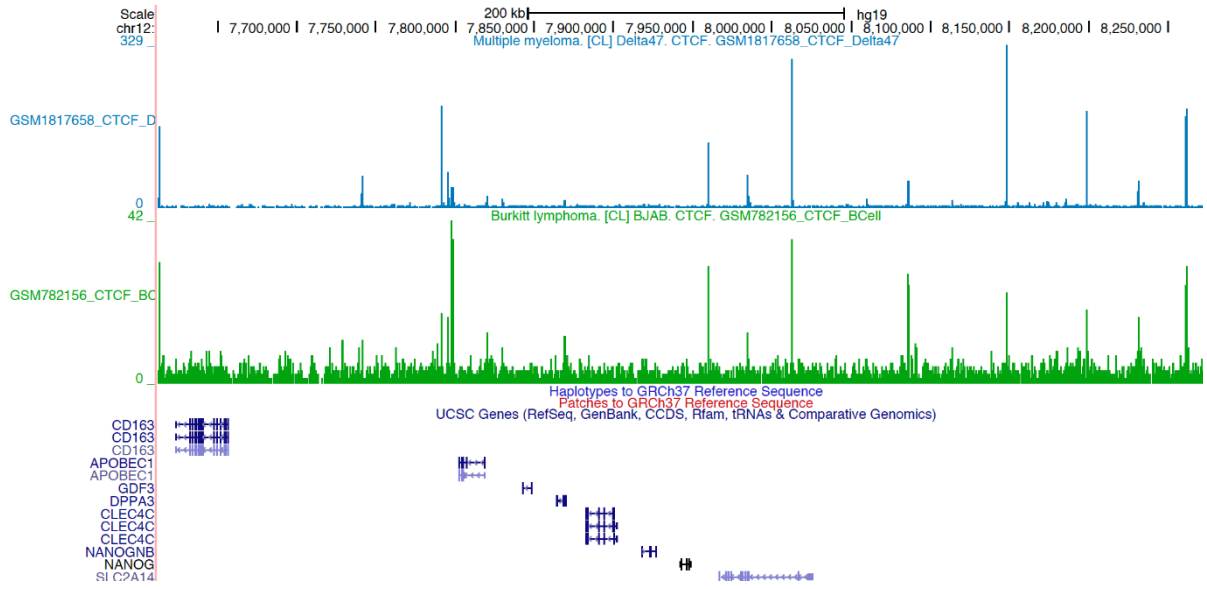


Figure 6.14 UCSC genome browser confirms similar binding profiles for CTCF at the *NANOG* locus for myeloma and Burkitt Lymphoma cell lines

One of the most useful features of CODEX is that it allows users to freely upload their own ChIP-seq data and compare it to existing CODEX profiles. For instance, Jiapaer *et al* used CODEX to identify that the promoter for long non-coding RNA LincU, which maintains embryonic stem cells in a naïve state, was bound by KLF4 and SMAD1 as well as NANOG, and then tested this transcriptional mechanism with a LincU-promoter luciferase reporter (Jiapaer *et al*, 2018).

In addition, CODEX can be employed to probe results from GWAS. For instance, if a SNP is identified which correlates with a clinical condition, then CODEX can inform users if that SNP is located at a locus with a particular TF or histone modification profile in a cell type of interest or across all cell types. Hodonsky *et al* performed a GWAS of 12,502 Hispanic people to characterise novel SNPs and genomic loci associated with red blood cell traits. They then probed this data for loci which were likely to be of functional significance, by cross-referencing with CODEX histone modification and ChIP-seq signals for key erythroid TFs (Hodonsky *et al*, 2017). Similarly, it can be informative to examine TF or histone binding profiles for mutations in non-coding regions of DNA. Another potential use of CODEX is to visualise whether altered gene expression in a specific human haematological malignancy correlates with TF binding to a particular locus.

In summary, our work has significantly expanded previous iterations of CODEX and we now have 459028795 bases with non-zero coverage, resulting in 14.8% of the genome covered with histone modification data. This latest iteration has significant potential for improving the applicability of CODEX to human disease states, as our focus has been to input data from human malignant blood cells. One of the strengths of CODEX is that its standardised pipeline permits continued expansion of data sets which will increase the power of pairwise correlations. In this regard, future work will include the incorporation of important cell types such as normal neutrophils and common malignant blood cells such as chronic lymphocytic leukaemia and follicular lymphoma. It will also be important to explore methods other than standard hierarchical clustering to identify putative groupings of experiments. Given the high-dimensional nature of the data contained in CODEX once several hundred experiments are considered, techniques now popular with single cell RNA-Seq analysis may be worth exploring, such as t-SNE and UMAP. It is also worth noting that CODEX appears to cope well with batch effects in general, since we repeatedly observe that experiments from different centres

demonstrate pairwise correlation. However, this is not always the case, suggesting that development of algorithms for batch effect removal may also be an important area for future investigations.

7. Conclusions

CEBPA double mutations confer favourable prognosis in patients with cytogenetically normal AML in the absence of *FLT3* ITD mutations, but it is important to remember that *CEBPA*-mutated AML nonetheless carries a significant mortality and morbidity burden, with a five-year overall survival of only 53% to 60% (Preudhomme *et al*, 2002, Renneville *et al*, 2009). Bearing in mind that *CEBPA* is one of the most common mutations implicated in human AML, the significant healthcare burden associated with treatment of this aggressive disease (Bewersdorf *et al*, 2019), and the fact that the mainstay of current chemotherapy regimens date from the 1970s (Evans *et al*, 1961), it can be persuasively argued that a better understanding of the biology of *CEBPA*-mutated AML is urgently required.

The cellular model for testing a specific mutational perturbation is particularly important when interpreting results of interventions. The impact of distinct models can be appreciated by the discordant results of *CEBPA* deletion on HSC function identified by Porse and colleagues who found that C/EBPα promotes HSC maintenance (Hasemann *et al*, 2014), as compared with Tenen and colleagues who employed a different model and concluded that C/EBPα inhibits HSC proliferation (Ye *et al*, 2013). Investigations of *CEBPA*-mutated AML have typically transduced myeloid-committed cells such as the 32Dcl3 cell line (Guchhait *et al*, 2003) for *in vitro* studies, and harvested cells from highly heterogeneous haematopoietic compartments such as bone marrow mononuclear cells (Togami *et al*, 2015) or foetal hepatocytes (Bereshchenko *et al*, 2009) for *in vivo* investigations. We have successfully created a novel *CEBPA*-mutated cell line which has several advantages over previous *in vitro* models. Firstly, our conditional model is based on a LMPP-like cell line which replicates a relatively early stage of haematopoiesis and is capable of reconstituting myeloid, lymphoid, and dendritic cell (DC) lineages. This is relevant because *CEBPA* is implicated as an early driver mutation in *de novo* AML, as evidenced by *CEBPA*-mutated clones being present in relapse, and we are more likely to recapitulate pre-leukaemic changes at the LMPP stage than in committed myeloid progenitors. Furthermore, several lines of investigation suggest that C/EBPα-mediated lineage specification is a relatively early event, for instance Welner *et al* found that C/EBPα was required for differentiation from HSPCs to common dendritic progenitors, but C/EBPα deletion in mature DCs had no effect on their numbers or function (Welner *et al*, 2013).

Secondly, the clonal murine Hoxb8-FL model provides a reproducible and homogenous cellular model which avoids the uncertainty associated with employing a highly heterogeneous cell compartment. Thirdly, we have employed the same model to generate Empty Vector and *CEBPA* WT-transduced cell lines to act as comparators.

The advantages of the precise control offered by this cellular model were initially evidenced by comparison of the behaviour of *CEBPA* WT-transduced cells in Hoxb8-FL conditions, which showed a cell marker profile consistent with differentiation. By contrast, *CEBPA* N321D-transduced cells failed to differentiate in Hoxb8-FL conditions, suggesting that the mutation has a dominant negative effect.

Our cellular model then allowed us to reproducibly test the effect of mutant C/EBPα on early differentiation in myeloid, lymphoid and DC lineages. We found that the N321D mutation had no appreciable effect on the cell marker profile of transduced cells in cytokine conditions which favoured granulocyte or lymphocyte specification, but showed a distinct phenotype in Flt3L, strongly suggesting an early DC effect in mutant C/EBPα. Furthermore, on prolonged cell culture, it was found that the C/EBPα N321D cell line was immortalised in Flt3L conditions.

Subsequent analysis by single cell RNA-seq (scRNA-seq) confirmed the C/EBPα WT and C/EBPα N321D phenotypes suggested by flow cytometry, and captured precisely the gene expression profile associated with the N321D mutation during early differentiation in Flt3L conditions. In addition, transcriptome analysis at the level of the single cell revealed a small sub-population of *CEBPA* N321D-transduced cells with a distinct gene expression profile, which would not have been captured by bulk RNA-seq and which can be hypothesised to represent the immortalised fraction of cells transduced with mutant *CEBPA*. This result underlined the importance of single cell analysis to appreciate the rare events implicated in heterogeneous pre-leukaemic alterations. ChIP-seq was then performed to identify putative direct binding events implicated in those alterations, and identified direct targets of C/EBPα N321D with altered gene expression, including *NOTCH2*, *JAK2*, *SIRPA* and *FOS* which perform significant roles in DC lineage specification and differentiation, as well as genes such as *MYB* and *E2F2* implicated in HSPC cell state and cell cycle control respectively. Interestingly, it was more difficult to identify a significant subset of these genes which had two-fold enrichment for H3K27Ac, and our results suggest that the N321D mutation does not obviously affect the

acetylation profile. This is a surprising result, given that biologically meaningful genes are bound by C/EBPα N321D and have altered gene expression in our experiments, and given that Pundhir *et al* have found that chromatin accessibility and activity correlate with C/EBPα's function as a pioneer factor in the context of early GM lineage differentiation (Pundhir *et al*, 2018). In this regard, it is important to remember that key genes such as *IRF8*, *RUNX1* and *SPI1* were bound by C/EBPα N321D but did not show altered gene expression, and it is tempting to speculate that the phenotypic effects of the N321D mutation may be mediated in part by conformational changes and altered dimerisation profiles rather than a direct binding event between C/EBPα N321D and *cis*-regulatory elements, explaining the disconnect between DNA binding patterns, acetylation profiles and gene expression. In this regard, Thomas Graf and colleagues have found that topological reorganisation often precedes gene expression changes during cell reprogramming by transient expression of *CEBPA* followed by induction of the Yamanaka TFs OCT4, SOX2, KLF4 and MYC (Stadhouders *et al*, 2018).

In mice transplanted with the C/EBPα N321D cell line, we found that mutant cells were capable of long-term engraftment *in vivo* but did not reliably reproduce leukaemia, suggesting that co-operating mutations are required for a leukaemic phenotype. This did not correlate with results reported by Togami *et al* when C57BL/6J mice received transplants of BM-derived cells tranduced with the N321D mutation (Togami *et al*, 2015). However, it does agree with reports by other groups including Nerlov and colleagues that C-terminal mutations result in expansion of multipotent progenitors with no progression to leukaemia (Bereshchenko *et al*, 2009). In this regard, the Bonnet group found long-term repopulating activity in human cord blood-derived Lin^- cells transduced with NC-terminal but not C-terminal mutated *CEBPA* (Quintana-Bustamente *et al*, 2012). Our results also correspond with several lines of clinical data from human patients with *CEBPA*-mutated AML:

(i) these patients are typically characterised by a bi-allelic mutational profile (Wouters *et al*, 2009);

(ii) hereditary germline mutations in *CEBPA* usually affect the N-terminus but do not result in a familial AML phenotype until patients acquire a somatic C-terminal mutation (Tawana *et al*, 2015);

(iii) *CEBPA* can cooperate with mutations in *ASXL1*, *CSF3R*, *FLT3*, *GATA2*, *RUNX1*, *TET2* and *WT1* (Fasan *et al*, 2014, Lavallee *et al*, 2016); and

(iv) AML patients with single-allele *CEBPA* mutations typically carry such a cooperating mutation (Wouters *et al*, 2009).

Recent work by van Galen *et al* has characterised heterogeneity in AML by performing scRNA-seq on 38,410 cells from sixteen AML patients and five healthy donors, and identified a spectrum of malignant cell types including cells with a DC transcriptional profile, as well as demonstrating that leukaemic cells have immunomodulatory properties such as T cell inhibition (van Galen *et al*, 2019). Our own experiments suggested that expression of *CEBPA* N321D at an early LMPP-like stage of haematopoiesis deregulates dendritic cell differentiation, specifically that it causes an immortalised $CD11c^{+}$ $B220^{+}$ $BST2^{-}$ pDC-like progenitor phenotype, correlating with gene set enrichment of differentially expressed genes identified by scRNA-seq. In addition, exogenous mutant C/EBPα binding events were increased in our cell line after differentiation, suggesting that the pre-leukaemic events occur in the HSPC compartment.

The role of C/EBPα in myeloid differentiation has been intensively studied, but relatively little is known regarding its function specifically in DC biology. Early work by Thomas Graf and colleagues suggested that expression of PU.1 in fully committed pre-T cells induces formation of dendritic cells, whereas expression of C/EBPα reprograms T cell progenitors to macrophages, and furthermore that NOTCH signalling is able to inhibit these processes (Laiosa *et al*, 2006). Interestingly, our experiments demonstrated that C/EBPα N321D binds directly with *NOTCH2* and that expression of *NOTCH2* is downregulated in *CEBPA* N321D-tranduced cells on differential expression with EV-transduced cells, and it is possible that NOTCH deregulation may influence DC lineage specification by the PU.1: C/EBPα axis. Later work by the Tenen group showed that there is an early requirement for C/EBPα in DC differentiation and that formation of mature DCs through myeloid progenitors is C/EBPα-dependent, whereas lymphoid progenitors are able to generate DCs in the absence of C/EBPα (Welner *et al*, 2013). Given that recent work by Rodrigues *et al* has shown that pDCs develop mostly from lymphoid progenitor cells (Rodrigues *et al*, 2018), it seems reasonable to hypothesise that the N321D mutation disrupts the C/EBPα-dependent myeloid pathway and favours the generation of pDCs rather than cDCs. This correlates with our experimental results that *CEBPA* N321D-mutated cells initially show reduced expression of cDC cell surface markers, before

progressing to an immortalised pDC-like immunophenotype. It is important to recognise that the phenotypic effects of C/EBPα N321D may also be mediated by interactions with other transcription factors implicated in lineage specification. For instance, our experiments have shown that mutant C/EBPα binds to *IRF8* and *SPI1*, though gene expression of these two genes was not directly altered.

Given that our work suggests for the first time that a pDC phenotype may characterise pre-leukaemic changes in mutant C/EBPα, it is therefore useful to consider our results within the broader context of DC biology. It is well-established that DC subsets include professional antigen presenting cells and immunomodulatory cells which secrete IFN-α. More recently, evidence suggests that pDCs have a unique TLR7-dependent capability to recognise virus-infected cells throughout their intracellular replication rather than only during their rare extracellular transit events (Takahashi *et al*, 2010), and this means that pDCs can counter effectively the multiple mechanisms of immune evasion deployed by viruses. Marlène Dreux and others have shown that sensing of infected cells by pDCs involves viral envelope protein-dependent secretion and transmission of viral RNA in cell-to-cell interactions (Decembre *et al*, 2014), and it is tempting to speculate that pDCs may employ a similar mechanism to play a unique role in cancer surveillance.

Regarding pDC function, Rodrigues *et al* have shown that lymphoid-pathway pDCs produce interferon but do not participate in antigen presentation (Rodrigues *et al*, 2018). It would certainly be interesting to perform functional evaluation of the *CEBPA* N321D immortalised pDC-like progenitor cell line, for instance by measuring IFN-α production upon TLR9 stimulation and T cell proliferation in response to LPS stimulation. Interestingly, T cell anergy has also been noted in AML-derived dendritic cells (Narita *et al*, 2001), and in a broader malignant context intratumoral pDCs have been functionally characterised in a murine mammary tumour model, favouring tumour progression by secreting IFN-α poorly and by inducing differentiation of regulatory T cells (Le Mercier *et al*, 2013). Besides the direct effects of the immortalised pDC progenitor, it is also important to bear in mind that one of our earliest experimental findings during culture in Flt3L differentiation media was that *CEBPA* N321D-transduced cells reduced expression of cDC type 2 markers CD11b and SIRPα. Conventional dendritic cells play an important role in cancer surveillance and presentation of tumour

antigen (Bottcher *et al*, 2018a, Bottcher *et al*, 2018b), and therefore N321D-mediated leukaemia may be favoured by dysregulation of the cDC compartment.

Further work is required to elucidate the effects of the N321D mutation in a complex biological system where cross-talk between IRF8, PU.1, C/EBPα and other TFs informs cell fate decisions between neutrophils, lymphocytes and mono-phagocytic cells. Regarding *in vitro* work, it would be important in future experiments to perform functional studies to confirm the immortalised pDC progenitor phenotype suggested by cell marker and gene expression profiles of *CEBPA* N321D-transduced cells. It would also be interesting to construct a more complex retroviral expression system where Hoxb8-FL cells are transduced with N and C-terminal mutations, to see if this recapitulates a more aggressive leukaemic phenotype. The intersection of scRNA-seq and ChIP-seq data has identified important putative genes implicated in pre-leukaemic changes, and ChIP-seq is particularly apt because of C/EBPα's role as a TF, but definitive therapeutic targets may be more clearly identified by CRISPR dropout screens (Tzelepis *et al*, 2016). In addition, scRNA-seq identified a very small subpopulation of five Hoxb8-FL cells transduced with mutant *CEBPA* which had a distinct "non-differentiated" transcriptional profile on t-SNE dimensionality reduction after five days of culture in Flt3L differentiation media. This subpopulation may represent leukaemia-initiating cells, and it would be interesting to employ 10x Chromium or DropSeq technology to process a larger number of single *CEBPA*-mutated Hoxb8-FL cells to capture more of these rare events. For *in vivo* studies, important future work includes (i) next generation sequencing of DNA from BM and splenic cells to characterise the full mutational profile of mice which developed latent leukaemia after transplantation of mutant cells, (ii) secondary and serial transplantation of *CEBPA* N321D-transduced cells to confirm their self-renewal capacity, and (iii) CRISPR-Cas9 genome editing to engineer *in vivo* models of cooperating mutations including *CEBPA* N321D (Heckl *et al*, 2014).

Lastly, this study opens up potential novel therapeutic options for the significant percentage of AML patients who carry *CEBPA* mutations. Previous targeted therapies for AML, and in particular APML, have focused on using CCAAT/Enhancer binding proteins to drive differentiation by inserting tamoxifen-responsive C/EBPε into a retroviral vector (Truong *et al*, 2003) or using 2-Cyano-3,12-dioxooleana-1,9-dien-28-oic acid to induce *de novo* synthesis of p42 C/EBPα (Koschmieder *et al*, 2007). Previous work has also identified specific molecular

targets for *CEBPA*-mutated AML, which include *SOX4* and *JAK* (Zhang *et al*, 2013, Lavallee *et al*, 2016). Our own experiments have combined an *in vitro* cellular model with genome-wide characterisation of DNA binding events, histone modification and gene expression at the level of the single cell to identify *NOTCH2*, *MYB* and other genes as putative targets for pharmacological intervention by small molecule activators (Ye *et al*, 2016) or inhibitors (Uttarkar *et al*, 2017) in *CEBPA*-mutated AML.

References

Abelson, S., G. Collord, S.W.K. Ng, O. Weissbrod, N. Mendelson Cohen, E. Niemeyer, N. Barda, P.C. Zuzarte, L. Heisler, Y. Sundaravadanam, R. Luben, S. Hayat, T.T. Wang, Z. Zhao, I. Cirlan, T.J. Pugh, D. Soave, K. Ng, C. Latimer, C. Hardy, K. Raine, D. Jones, D. Hoult, A. Britten, J.D. McPherson, M. Johansson, F. Mbabaali, J. Eagles, J.K. Miller, D. Pasternack, L. Timms, P. Krzyzanowski, P. Awadalla, R. Costa, E. Segal, S.V. Bratman, P. Beer, S. Behjati, I. Martincorena, J.C.Y. Wang, K.M. Bowles, J.R. Quiros, A. Karakatsani, C. La Vecchia, A. Trichopoulou, E. Salamanca-Fernandez, J.M. Huerta, A. Barricarte, R.C. Travis, R. Tumino, G. Masala, H. Boeing, S. Panico, R. Kaaks, A. Kramer, S. Sieri, E. Riboli, P. Vineis, M. Foll, J. McKay, S. Polidoro, N. Sala, K.T. Khaw, R. Vermeulen, P.J. Campbell, E. Papaemmanuil, M.D. Minden, A. Tanay, R.D. Balicer, N.J. Wareham, M. Gerstung, J.E. Dick, P. Brennan, G.S. Vassiliou, and L.I. Shlush. 2018. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature*. 559:400-404.

Adelman, E.R., H.T. Huang, A. Roisman, A. Olsson, A. Colaprico, T. Qin, R.C. Lindsley, R. Bejar, N. Salomonis, H.L. Grimes, and M.E. Figueroa. 2019. Aging Human Hematopoietic Stem Cells Manifest Profound Epigenetic Reprogramming of Enhancers That May Predispose to Leukemia. *Cancer Discov*. 9:1080-1101.

Adolfsson, J., R. Mansson, N. Buza-Vidas, A. Hultquist, K. Liuba, C.T. Jensen, D. Bryder, L. Yang, O.J. Borge, L.A. Thoren, K. Anderson, E. Sitnicka, Y. Sasaki, M. Sigvardsson, and S.E. Jacobsen. 2005. Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell*. 121:295-306.

Ahmed, M.S., S.E. Byeon, Y. Jeong, M.A. Miah, M. Salahuddin, Y. Lee, S.S. Park, and Y.S. Bae. 2015. Dab2, a negative regulator of DC immunogenicity, is an attractive molecular target for DC-based immunotherapy. *Oncoimmunology*. 4:e984550.

Ahn, J.H., Y. Lee, C. Jeon, S.J. Lee, B.H. Lee, K.D. Choi, and Y.S. Bae. 2002. Identification of the genes differentially expressed in human dendritic cell subsets by cDNA subtraction and microarray analysis. *Blood*. 100:1742-1754.

Akashi, K., D. Traver, T. Miyamoto, and I.L. Weissman. 2000. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*. 404:193-197.

Allman, D., M. Dalod, C. Asselin-Paturel, T. Delale, S.H. Robbins, G. Trinchieri, C.A. Biron, P. Kastner, and S. Chan. 2006. Ikaros is required for plasmacytoid dendritic cell differentiation. *Blood*. 108:4025-4034.

Aoki, N., Y. Kimura, S. Kimura, T. Nagato, M. Azumi, H. Kobayashi, K. Sato, and M. Tateno. 2009. Expression and functional role of MDL-1 (CLEC5A) in mouse myeloid lineage cells. *J Leukoc Biol*. 85:508-517.

Aparicio, S., and C. Caldas. 2013. The implications of clonal genome evolution for cancer medicine. *N Engl J Med*. 368:842-851.

Arber, D.A., A. Orazi, R. Hasserjian, J. Thiele, M.J. Borowitz, M.M. Le Beau, C.D. Bloomfield, M. Cazzola, and J.W. Vardiman. 2016. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 127:2391-2405.

Asselin-Paturel, C., A. Boonstra, M. Dalod, I. Durand, N. Yessaad, C. Dezutter-Dambuyant, A. Vicari, A. O'Garra, C. Biron, F. Briere, and G. Trinchieri. 2001. Mouse type I IFN-producing cells are immature APCs with plasmacytoid morphology. *Nat Immunol*. 2:1144-1150.

Avellino, R., M. Havermans, C. Erpelinck, M.A. Sanders, R. Hoogenboezem, H.J. van de Werken, E. Rombouts, K. van Lom, P.M. van Strien, C. Gebhard, M. Rehli, J. Pimanda, D. Beck, S. Erkeland, T. Kuiken, H. de Looper, S. Groschel, I. Touw, E. Bindels, and R. Delwel. 2016. An autonomous CEBPA enhancer specific for myeloid-lineage priming and neutrophilic differentiation. *Blood*.

127:2991-3003.

Bachem, A., E. Hartung, S. Guttler, A. Mora, X. Zhou, A. Hegemann, M. Plantinga, E. Mazzini, P. Stoitzner, S. Gurka, V. Henn, H.W. Mages, and R.A. Kroczek. 2012. Expression of XCR1 Characterizes the Batf3-Dependent Lineage of Dendritic Cells Capable of Antigen Cross-Presentation. *Front Immunol*. 3:214.

Baran-Gale, J., T. Chandra, and K. Kirschner. 2018. Experimental design for single-cell RNA sequencing. *Brief Funct Genomics*. 17:233-239.

Baron, C.S., L. Kester, A. Klaus, J.C. Boisset, R. Thambyrajah, L. Yvernogeau, V. Kouskoff, G. Lacaud, A. van Oudenaarden, and C. Robin. 2018. Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the aorta. *Nat Commun*. 9:2517.

Bartenhagen, C., H.U. Klein, C. Ruckert, X. Jiang, and M. Dugas. 2010. Comparative study of unsupervised dimension reduction techniques for the visualization of microarray gene expression data. *BMC Bioinformatics*. 11:567.

Bartman, C.R., S.C. Hsu, C.C. Hsiung, A. Raj, and G.A. Blobel. 2016. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol Cell*. 62:237-247.

Basso, K., A.A. Margolin, G. Stolovitzky, U. Klein, R. Dalla-Favera, and A. Califano. 2005. Reverse engineering of regulatory networks in human B cells. *Nat Genet*. 37:382-390.

Batliner, J., M.M. Mancarelli, M. Jenal, V.A. Reddy, M.F. Fey, B.E. Torbett, and M.P. Tschan. 2011. CLEC5A (MDL-1) is a novel PU.1 transcriptional target during myeloid differentiation. *Mol Immunol*. 48:714-719.

Beck, D., J.A. Thoms, D. Perera, J. Schutte, A. Unnikrishnan, K. Knezevic, S.J. Kinston, N.K. Wilson, T.A. O'Brien, B. Gottgens, J.W. Wong, and J.E. Pimanda. 2013. Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood*. 122:e12-22.

Becker, A.J., C.E. Mc, and J.E. Till. 1963. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature*. 197:452-454.

Belluschi, S., E.F. Calderbank, V. Ciaurro, B. Pijuan-Sala, A. Santoro, N. Mende, E. Diamanti, K.Y.C. Sham, X. Wang, W.W.Y. Lau, W. Jawaid, B. Gottgens, and E. Laurenti. 2018. Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. *Nat Commun*. 9:4100.

Belz, G.T., K. Shortman, M.J. Bevan, and W.R. Heath. 2005. CD8alpha+ dendritic cells selectively present MHC class I-restricted noncytolytic viral and intracellular bacterial antigens in vivo. *J Immunol*. 175:196-200.

Bene, M.C., G. Castoldi, W. Knapp, W.D. Ludwig, E. Matutes, A. Orfao, and M.B. van't Veer. 1995. Proposals for the immunological classification of acute leukemias. European Group for the Immunological Characterization of Leukemias (EGIL). *Leukemia*. 9:1783-1786.

Bengtsson, M., A. Stahlberg, P. Rorsman, and M. Kubista. 2005. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Res*. 15:1388-1392.

Bennett, J.M., D. Catovsky, M.T. Daniel, G. Flandrin, D.A. Galton, H.R. Gralnick, and C. Sultan. 1976. Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol*. 33:451-458.

Bereshchenko, O., E. Mancini, S. Moore, D. Bilbao, R. Mansson, S. Luc, A. Grover, S.E. Jacobsen, D. Bryder, and C. Nerlov. 2009. Hematopoietic stem cell expansion precedes the generation of committed myeloid leukemia-initiating cells in C/EBPalpha mutant AML. *Cancer Cell*. 16:390-400.

Bewersdorf, J.P., R.M. Shallis, R. Wang, S.F. Huntington, S. Perreault, X. Ma, and A.M. Zeidan. 2019. Healthcare expenses for treatment of acute myeloid leukemia. *Expert Rev Hematol*. 12:641-650.

Bonnet, D., and J.E. Dick. 1997. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med*. 3:730-737.

Bottcher, J.P., E. Bonavita, P. Chakravarty, H. Blees, M. Cabeza-Cabrerizo, S. Sammicheli, N.C. Rogers, E. Sahai, S. Zelenay, and C. Reis e Sousa. 2018. NK Cells Stimulate Recruitment of cDC1 into the Tumor Microenvironment Promoting Cancer Immune Control. *Cell*. 172:1022-1037 e1014.

Bottcher, J.P., and E.S.C. Reis. 2018. The Role of Type 1 Conventional Dendritic Cells in Cancer Immunity. *Trends Cancer*. 4:784-792.

Braun, C.J., K. Boztug, A. Paruzynski, M. Witzel, A. Schwarzer, M. Rothe, U. Modlich, R. Beier, G. Gohring, D. Steinemann, R. Fronza, C.R. Ball, R. Haemmerle, S. Naundorf, K. Kuhlcke, M. Rose, C. Fraser, L. Mathias, R. Ferrari, M.R. Abboud, W. Al-Herz, I. Kondratenko, L. Marodi, H. Glimm, B. Schlegelberger, A. Schambach, M.H. Albert, M. Schmidt, C. von Kalle, and C. Klein. 2014. Gene therapy for Wiskott-Aldrich syndrome--long-term efficacy and genotoxicity. *Sci Transl Med*. 6:227ra233.

Brennecke, P., S. Anders, J.K. Kim, A.A. Kolodziejczyk, X. Zhang, V. Proserpio, B. Baying, V. Benes, S.A. Teichmann, J.C. Marioni, and M.G. Heisler. 2013. Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods*. 10:1093-1095.

Brunetti, L., M.C. Gundry, D. Sorcini, A.G. Guzman, Y.H. Huang, R. Ramabadran, I. Gionfriddo, F. Mezzasoma, F. Milano, B. Nabet, D.L. Buckley, S.M. Kornblau, C.Y. Lin, P. Sportoletti, M.P. Martelli, B. Falini, and M.A. Goodell. 2018. Mutant NPM1 Maintains the Leukemic State through HOX Expression. *Cancer Cell*. 34:499-512 e499.

Buenrostro, J.D., M.R. Corces, C.A. Lareau, B. Wu, A.N. Schep, M.J. Aryee, R. Majeti, H.Y. Chang, and W.J. Greenleaf. 2018. Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell*. 173:1535-1548 e1516.

Buenrostro, J.D., B. Wu, U.M. Litzenburger, D. Ruff, M.L. Gonzales, M.P. Snyder, H.Y. Chang, and W.J. Greenleaf. 2015. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*. 523:486-490.

Buettner, F., K.N. Natarajan, F.P. Casale, V. Proserpio, A. Scialdone, F.J. Theis, S.A. Teichmann, J.C. Marioni, and O. Stegle. 2015. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol*. 33:155-160.

Burger, J.A., and A. Burkle. 2007. The CXCR4 chemokine receptor in acute and chronic leukaemia: a marrow homing receptor and potential therapeutic target. *Br J Haematol*. 137:288-296.

Cai, D.H., D. Wang, J. Keefer, C. Yeamans, K. Hensley, and A.D. Friedman. 2008. C/EBP alpha:AP-1 leucine zipper heterodimers bind novel DNA elements, activate the PU.1 promoter and direct monocyte lineage commitment more potently than C/EBP alpha homodimers or AP-1. *Oncogene*. 27:2772-2779.

Calero-Nieto, F.J., A.G. Bert, and P.N. Cockerill. 2010. Transcription-dependent silencing of inducible convergent transgenes in transgenic mice. *Epigenetics Chromatin*. 3:3.

Calero-Nieto, F.J., F.S. Ng, N.K. Wilson, R. Hannah, V. Moignard, A.I. Leal-Cervantes, I. Jimenez-Madrid, E. Diamanti, L. Wernisch, and B. Gottgens. 2014. Key regulators control distinct transcriptional programmes in blood progenitor and mast cells. *EMBO J*. 33:1212-1226.

Calvi, L.M., G.B. Adams, K.W. Weibrecht, J.M. Weber, D.P. Olson, M.C. Knight, R.P. Martin, E. Schipani, P. Divieti, F.R. Bringhurst, L.A. Milner, H.M. Kronenberg, and D.T. Scadden. 2003. Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature*. 425:841-846.

Cancer Genome Atlas Research Network, T.J. Ley, C. Miller, L. Ding, B.J. Raphael, A.J. Mungall, A.

Robertson, K. Hoadley, T.J. Triche, Jr., P.W. Laird, J.D. Baty, L.L. Fulton, R. Fulton, S.E. Heath, J. Kalicki-Veizer, C. Kandoth, J.M. Klco, D.C. Koboldt, K.L. Kanchi, S. Kulkarni, T.L. Lamprecht, D.E. Larson, L. Lin, C. Lu, M.D. McLellan, J.F. McMichael, J. Payton, H. Schmidt, D.H. Spencer, M.H. Tomasson, J.W. Wallis, L.D. Wartman, M.A. Watson, J. Welch, M.C. Wendl, A. Ally, M. Balasundaram, I. Birol, Y. Butterfield, R. Chiu, A. Chu, E. Chuah, H.J. Chun, R. Corbett, N. Dhalla, R. Guin, A. He, C. Hirst, M. Hirst, R.A. Holt, S. Jones, A. Karsan, D. Lee, H.I. Li, M.A. Marra, M. Mayo, R.A. Moore, K. Mungall, J. Parker, E. Pleasance, P. Plettner, J. Schein, D. Stoll, L. Swanson, A. Tam, N. Thiessen, R. Varhol, N. Wye, Y. Zhao, S. Gabriel, G. Getz, C. Sougnez, L. Zou, M.D. Leiserson, F. Vandin, H.T. Wu, F. Applebaum, S.B. Baylin, R. Akbani, B.M. Broom, K. Chen, T.C. Motter, K. Nguyen, J.N. Weinstein, N. Zhang, M.L. Ferguson, C. Adams, A. Black, J. Bowen, J. Gastier-Foster, T. Grossman, T. Lichtenberg, L. Wise, T. Davidsen, J.A. Demchok, K.R. Shaw, M. Sheth, H.J. Sofia, L. Yang, J.R. Downing, et al. 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 368:2059-2074.

Cao, W., A.K. Taylor, R.E. Biber, W.G. Davis, J.H. Kim, A.J. Reber, T. Chirkova, J.A. De La Cruz, A. Pandey, P. Ranjan, J.M. Katz, S. Gangappa, and S. Sambhara. 2012. Rapid differentiation of monocytes into type I IFN-producing myeloid dendritic cells as an antiviral strategy against influenza virus infection. *J Immunol*. 189:2257-2265.

Cao, Z., R.M. Umek, and S.L. McKnight. 1991. Regulated expression of three C/EBP isoforms during adipose conversion of 3T3-L1 cells. *Genes Dev*. 5:1538-1552.

Carotta, S., A. Dakic, A. D'Amico, S.H. Pang, K.T. Greig, S.L. Nutt, and L. Wu. 2010. The transcription factor PU.1 controls dendritic cell development and Flt3 cytokine receptor expression in a dose-dependent manner. *Immunity*. 32:628-641.

Challen, G.A., D. Sun, M. Jeong, M. Luo, J. Jelinek, J.S. Berg, C. Bock, A. Vasanthakumar, H. Gu, Y. Xi, S. Liang, Y. Lu, G.J. Darlington, A. Meissner, J.P. Issa, L.A. Godley, W. Li, and M.A. Goodell. 2011. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet*. 44:23-31.

Chamorro, S., J.J. Garcia-Vallejo, W.W. Unger, R.J. Fernandes, S.C. Bruijns, S. Laban, B.O. Roep, B.A. t Hart, and Y. van Kooyk. 2009. TLR triggering on tolerogenic dendritic cells results in TLR2 up-regulation and a reduced proinflammatory immune program. *J Immunol*. 183:2984-2994.

Charbord, P., C. Pouget, H. Binder, F. Dumont, G. Stik, P. Levy, F. Allain, C. Marchal, J. Richter, B. Uzan, F. Pflumio, F. Letourneur, H. Wirth, E. Dzierzak, D. Traver, T. Jaffredo, and C. Durand. 2014. A systems biology approach for defining the molecular framework of the hematopoietic stem cell niche. *Cell Stem Cell*. 15:376-391.

Chen, E.Y., C.M. Tan, Y. Kou, Q. Duan, Z. Wang, G.V. Meirelles, N.R. Clark, and A. Ma'ayan. 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 14:128.

Chen, H., Y. Tian, W. Shu, X. Bo, and S. Wang. 2012. Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS One*. 7:e41374.

Cheng, P., V. Kumar, H. Liu, J.I. Youn, M. Fishman, S. Sherman, and D. Gabrilovich. 2014. Effects of notch signaling on regulation of myeloid cell differentiation in cancer. *Cancer Res*. 74:141-152.

Cheng, P., Y. Nefedova, C.A. Corzo, and D.I. Gabrilovich. 2007. Regulation of dendritic-cell differentiation by bone marrow stroma via different Notch ligands. *Blood*. 109:507-515.

Chitteti, B.R., M.A. Kacena, S.L. Voytik-Harbin, and E.F. Srour. 2015. Modulation of hematopoietic progenitor cell fate in vitro by varying collagen oligomer matrix stiffness in the presence or absence of osteoblasts. *J Immunol Methods*. 425:108-113.

Chopin, M., S.P. Preston, A.T.L. Lun, J. Tellier, G.K. Smyth, M. Pellegrini, G.T. Belz, L.M. Corcoran, J.E. Visvader, L. Wu, and S.L. Nutt. 2016. RUNX2 Mediates Plasmacytoid Dendritic Cell Egress from the Bone Marrow and Controls Viral Immunity. *Cell Rep*. 15:866-878.

Clark, G.J., X. Ju, M. Azlan, C. Tate, Y. Ding, and D.N. Hart. 2009. The CD300 molecules regulate monocyte and dendritic cell functions. *Immunobiology*. 214:730-736.

Cleaves, R., Q.F. Wang, and A.D. Friedman. 2004. C/EBPalphap30, a myeloid leukemia oncoprotein, limits G-CSF receptor expression but not terminal granulopoiesis via site-selective inhibition of C/EBP DNA binding. *Oncogene*. 23:716-725.

Collins, C., J. Wang, H. Miao, J. Bronstein, H. Nawer, T. Xu, M. Figueroa, A.G. Muntean, and J.L. Hess. 2014. C/EBPalpha is an essential collaborator in Hoxa9/Meis1-mediated leukemogenesis. *Proc Natl Acad Sci U S A*. 111:9899-9904.

Cooper, S., H. Guo, and A.D. Friedman. 2015. The +37 kb Cebpa Enhancer Is Critical for Cebpa Myeloid Gene Expression and Contains Functional Sites that Bind SCL, GATA2, C/EBPalpha, PU.1, and Additional Ets Factors. *PLoS One*. 10:e0126385.

Cortes, J.E., S. Khaled, G. Martinelli, A.E. Perl, S. Ganguly, N. Russell, A. Kramer, H. Dombret, D. Hogge, B.A. Jonas, A.Y. Leung, P. Mehta, P. Montesinos, M. Radsak, S. Sica, M. Arunachalam, M. Holmes, K. Kobayashi, R. Namuyinga, N. Ge, A. Yver, Y. Zhang, and M.J. Levis. 2019. Quizartinib versus salvage chemotherapy in relapsed or refractory FLT3-ITD acute myeloid leukaemia (QuANTUM-R): a multicentre, randomised, controlled, open-label, phase 3 trial. *Lancet Oncol*. 20:984-997.

Cozzio, A., E. Passegue, P.M. Ayton, H. Karsunky, M.L. Cleary, and I.L. Weissman. 2003. Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. *Genes Dev*. 17:3029-3035.

Creyghton, M.P., A.W. Cheng, G.G. Welstead, T. Kooistra, B.W. Carey, E.J. Steine, J. Hanna, M.A. Lodato, G.M. Frampton, P.A. Sharp, L.A. Boyer, R.A. Young, and R. Jaenisch. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*. 107:21931-21936.

Dahl, R., J.C. Walsh, D. Lancki, P. Laslo, S.R. Iyer, H. Singh, and M.C. Simon. 2003. Regulation of macrophage and neutrophil cell fates by the PU.1:C/EBPalpha ratio and granulocyte colony-stimulating factor. *Nat Immunol*. 4:1029-1036.

Dahlin, J.S., F.K. Hamey, B. Pijuan-Sala, M. Shepherd, W.W.Y. Lau, S. Nestorowa, C. Weinreb, S. Wolock, R. Hannah, E. Diamanti, D.G. Kent, B. Gottgens, and N.K. Wilson. 2018. A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. *Blood*. 131:e1-e11.

D'Amico, A., and L. Wu. 2003. The early progenitors of mouse dendritic cells and plasmacytoid predendritic cells are within the bone marrow hemopoietic precursors expressing Flt3. *J Exp Med*. 198:293-303.

Davis, R.L., H. Weintraub, and A.B. Lassar. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*. 51:987-1000.

Decembre, E., S. Assil, M.L. Hillaire, W. Dejnirattisai, J. Mongkolsapaya, G.R. Screaton, A.D. Davidson, and M. Dreux. 2014. Sensing of immature particles produced by dengue virus infected cells induces an antiviral response by plasmacytoid dendritic cells. *PLoS Pathog*. 10:e1004434.

Devailly, G., A. Mantsoki, and A. Joshi. 2016. Heat*seq: an interactive web tool for high-throughput sequencing experiment comparison with public data. *Bioinformatics*. 32:3354-3356.

Dey, S.S., J.E. Foley, P. Limsirichai, D.V. Schaffer, and A.P. Arkin. 2015. Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. *Mol Syst Biol*. 11:806.

Dhodapkar, M.V., R.M. Steinman, M. Sapp, H. Desai, C. Fossella, J. Krasovsky, S.M. Donahoe, P.R. Dunbar, V. Cerundolo, D.F. Nixon, and N. Bhardwaj. 1999. Rapid generation of broad T-cell immunity in humans after a single injection of mature dendritic cells. *J Clin Invest*. 104:173-180.

Di Stefano, B., S. Collombet, J.S. Jakobsen, M. Wierer, J.L. Sardina, A. Lackner, R. Stadhouders, C. Segura-Morales, M. Francesconi, F. Limone, M. Mann, B. Porse, D. Thieffry, and T. Graf. 2016. C/EBPalpha creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4. *Nat Cell Biol*. 18:371-381.

Di Stefano, B., J.L. Sardina, C. van Oevelen, S. Collombet, E.M. Kallin, G.P. Vicent, J. Lu, D. Thieffry, M. Beato, and T. Graf. 2014. C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature*. 506:235-239.

Diamond, M.S., M. Kinder, H. Matsushita, M. Mashayekhi, G.P. Dunn, J.M. Archambault, H. Lee, C.D. Arthur, J.M. White, U. Kalinke, K.M. Murphy, and R.D. Schreiber. 2011. Type I interferon is selectively required by dendritic cells for immune rejection of tumors. *J Exp Med*. 208:1989-2003.

Dickson, G.J., S. Bustraan, R.K. Hills, A. Ali, A.H. Goldstone, A.K. Burnett, D.C. Linch, and R.E. Gale. 2016. The value of molecular stratification for CEBPA(DM) and NPM1(MUT) FLT3(WT) genotypes in older patients with acute myeloid leukaemia. *Br J Haematol*. 172:573-580.

Ding, L., T.J. Ley, D.E. Larson, C.A. Miller, D.C. Koboldt, J.S. Welch, J.K. Ritchey, M.A. Young, T. Lamprecht, M.D. McLellan, J.F. McMichael, J.W. Wallis, C. Lu, D. Shen, C.C. Harris, D.J. Dooling, R.S. Fulton, L.L. Fulton, K. Chen, H. Schmidt, J. Kalicki-Veizer, V.J. Magrini, L. Cook, S.D. McGrath, T.L. Vickery, M.C. Wendl, S. Heath, M.A. Watson, D.C. Link, M.H. Tomasson, W.D. Shannon, J.E. Payton, S. Kulkarni, P. Westervelt, M.J. Walter, T.A. Graubert, E.R. Mardis, R.K. Wilson, and J.F. DiPersio. 2012. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*. 481:506-510.

Dixit, A., O. Parnas, B. Li, J. Chen, C.P. Fulco, L. Jerby-Arnon, N.D. Marjanovic, D. Dionne, T. Burks, R. Raychowdhury, B. Adamson, T.M. Norman, E.S. Lander, J.S. Weissman, N. Friedman, and A. Regev. 2016. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*. 167:1853-1866 e1817.

Doulatov, S., F. Notta, K. Eppert, L.T. Nguyen, P.S. Ohashi, and J.E. Dick. 2010. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat Immunol*. 11:585-593.

Duarte, D., S. Amarteifio, H. Ang, I.Y. Kong, N. Ruivo, G. Pruessner, E.D. Hawkins, and C. Lo Celso. 2019. Defining the in vivo characteristics of acute myeloid leukemia cells behavior by intravital imaging. *Immunol Cell Biol*. 97:229-235.

Edwards, D.K.t., K. Watanabe-Smith, A. Rofelty, A. Damnernsawad, T. Laderas, A. Lamble, E.F. Lind, A. Kaempf, M. Mori, M. Rosenberg, A. d'Almeida, N. Long, A. Agarwal, D.T. Sweeney, M. Loriaux, S.K. McWeeney, and J.W. Tyner. 2019. CSF1R inhibitors exhibit antitumor activity in acute myeloid leukemia by blocking paracrine signals from support cells. *Blood*. 133:588-599.

Evans, J.S., E.A. Musser, G.D. Mengel, K.R. Forsblad, and J.H. Hunter. 1961. Antitumor activity of 1-beta-D-arainofuranosylcytosine hydrochloride. *Proc Soc Exp Biol Med*. 106:350-353.

Fasan, A., C. Haferlach, T. Alpermann, S. Jeromin, V. Grossmann, C. Eder, S. Weissmann, F. Dicker, A. Kohlmann, S. Schindela, W. Kern, T. Haferlach, and S. Schnittger. 2014. The role of different genetic subtypes of CEBPA mutated AML. *Leukemia*. 28:794-803.

Federzoni, E.A., M. Humbert, B.E. Torbett, G. Behre, M.F. Fey, and M.P. Tschan. 2014. CEBPA-dependent HK3 and KLF5 expression in primary AML and during AML differentiation. *Sci Rep*. 4:4261.

Feinberg, M.W., A.K. Wara, Z. Cao, M.A. Lebedeva, F. Rosenbauer, H. Iwasaki, H. Hirai, J.P. Katz, R.L. Haspel, S. Gray, K. Akashi, J. Segre, K.H. Kaestner, D.G. Tenen, and M.K. Jain. 2007. The Kruppel-like factor KLF4 is a critical regulator of monocyte differentiation. *EMBO J*. 26:4138-4148.

Feng, R., S.C. Desbordes, H. Xie, E.S. Tillo, F. Pixley, E.R. Stanley, and T. Graf. 2008. PU.1 and

C/EBPalpha/beta convert fibroblasts into macrophage-like cells. *Proc Natl Acad Sci U S A*. 105:6057-6062.

Fischbach, N.A., S. Rozenfeld, W. Shen, S. Fong, D. Chrobak, D. Ginzinger, S.C. Kogan, A. Radhakrishnan, M.M. Le Beau, C. Largman, and H.J. Lawrence. 2005. HOXB6 overexpression in murine bone marrow immortalizes a myelomonocytic precursor in vitro and causes hematopoietic stem cell expansion and acute myeloid leukemia in vivo. *Blood*. 105:1456-1466.

Fogg, D.K., C. Sibon, C. Miled, S. Jung, P. Aucouturier, D.R. Littman, A. Cumano, and F. Geissmann. 2006. A clonogenic bone marrow progenitor specific for macrophages and dendritic cells. *Science*. 311:83-87.

Ford, A.M., S.A. Ridge, M.E. Cabrera, H. Mahmoud, C.M. Steel, L.C. Chan, and M. Greaves. 1993. In utero rearrangements in the trithorax-related oncogene in infant leukaemias. *Nature*. 363:358-360.

Forsberg, E.C., S.S. Prohaska, S. Katzman, G.C. Heffner, J.M. Stuart, and I.L. Weissman. 2005. Differential expression of novel potential regulators in hematopoietic stem cells. *PLoS Genet*. 1:e28.

Friedman, A.D. 2015. C/EBPalpha in normal and malignant myelopoiesis. *Int J Hematol*. 101:330-341.

Frolov, M.V., D.S. Huen, O. Stevaux, D. Dimova, K. Balczarek-Strang, M. Elsdon, and N.J. Dyson. 2001. Functional antagonism between E2F family members. *Genes Dev*. 15:2146-2160.

Gardeux, V., F.P.A. David, A. Shajkofci, P.C. Schwalie, and B. Deplancke. 2017. ASAP: a web-based platform for the analysis and interactive visualization of single-cell RNA-seq data. *Bioinformatics*. 33:3123-3125.

Gilliland, D.G., and J.D. Griffin. 2002. The roles of FLT3 in hematopoiesis and leukemia. *Blood*. 100:1532-1542.

Giustacchini, A., S. Thongjuea, N. Barkas, P.S. Woll, B.J. Povinelli, C.A.G. Booth, P. Sopp, R. Norfo, A. Rodriguez-Meira, N. Ashley, L. Jamieson, P. Vyas, K. Anderson, A. Segerstolpe, H. Qian, U. Olsson-Stromberg, S. Mustjoki, R. Sandberg, S.E.W. Jacobsen, and A.J. Mead. 2017. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat Med*. 23:692-702.

Goardon, N., E. Marchi, A. Atzberger, L. Quek, A. Schuh, S. Soneji, P. Woll, A. Mead, K.A. Alford, R. Rout, S. Chaudhury, A. Gilkes, S. Knapper, K. Beldjord, S. Begum, S. Rose, N. Geddes, M. Griffiths, G. Standen, A. Sternberg, J. Cavenagh, H. Hunter, D. Bowen, S. Killick, L. Robinson, A. Price, E. Macintyre, P. Virgo, A. Burnett, C. Craddock, T. Enver, S.E. Jacobsen, C. Porcher, and P. Vyas. 2011. Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell*. 19:138-152.

Gomes, A.M., I. Kurochkin, B. Chang, M. Daniel, K. Law, N. Satija, A. Lachmann, Z. Wang, L. Ferreira, A. Ma'ayan, B.K. Chen, D. Papatsenko, I.R. Lemischka, K.A. Moore, and C.F. Pereira. 2018. Cooperative Transcription Factor Induction Mediates Hemogenic Reprogramming. *Cell Rep*. 25:2821-2835 e2827.

Gou, H., J. Zhou, Y. Ye, X. Hu, M. Shang, J. Zhang, Z. Zhao, W. Peng, Y. Zhou, Y. Zhou, X. Song, X. Lu, and B. Ying. 2016. The prevalence and clinical profiles of FLT3-ITD, FLT3-TKD, NPM1, C-KIT, DNMT3A, and CEBPA mutations in a cohort of patients with de novo acute myeloid leukemia from southwest China. *Tumour Biol*. 37:7357-7370.

Graf, T., and T. Enver. 2009. Forcing cells to change lineages. *Nature*. 462:587-594.

Graham, F.L., J. Smiley, W.C. Russell, and R. Nairn. 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol*. 36:59-74.

Grajales-Reyes, G.E., A. Iwata, J. Albring, X. Wu, R. Tussiwand, W. Kc, N.M. Kretzer, C.G. Briseno, V. Durai, P. Bagadia, M. Haldar, J. Schonheit, F. Rosenbauer, T.L. Murphy, and K.M. Murphy. 2015.

Batf3 maintains autoactivation of Irf8 for commitment of a CD8alpha(+) conventional DC clonogenic progenitor. *Nat Immunol*. 16:708-717.

Grigoryan, G., and A.E. Keating. 2008. Structural specificity in coiled-coil interactions. *Curr Opin Struct Biol*. 18:477-483.

Grimwade, D., H. Walker, F. Oliver, K. Wheatley, C. Harrison, G. Harrison, J. Rees, I. Hann, R. Stevens, A. Burnett, and A. Goldstone. 1998. The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood*. 92:2322-2333.

Grouard, G., M.C. Rissoan, L. Filgueira, I. Durand, J. Banchereau, and Y.J. Liu. 1997. The enigmatic plasmacytoid T cells develop into dendritic cells with interleukin (IL)-3 and CD40-ligand. *J Exp Med*. 185:1101-1111.

Guchhait, P., M.F. Tosi, C.W. Smith, and A. Chakaraborty. 2003. The murine myeloid cell line 32Dcl3 as a model system for studying neutrophil functions. *J Immunol Methods*. 283:195-204.

Guermonprez, P., J. Valladeau, L. Zitvogel, C. Thery, and S. Amigorena. 2002. Antigen presentation and T cell stimulation by dendritic cells. *Annu Rev Immunol*. 20:621-667.

Guilliams, M., C.A. Dutertre, C.L. Scott, N. McGovern, D. Sichien, S. Chakarov, S. Van Gassen, J. Chen, M. Poidinger, S. De Prijck, S.J. Tavernier, I. Low, S.E. Irac, C.N. Mattar, H.R. Sumatoh, G.H.L. Low, T.J.K. Chung, D.K.H. Chan, K.K. Tan, T.L.K. Hon, E. Fossum, B. Bogen, M. Choolani, J.K.Y. Chan, A. Larbi, H. Luche, S. Henri, Y. Saeys, E.W. Newell, B.N. Lambrecht, B. Malissen, and F. Ginhoux. 2016. Unsupervised High-Dimensional Analysis Aligns Dendritic Cells across Tissues and Species. *Immunity*. 45:669-684.

Guo, H., S. Cooper, and A.D. Friedman. 2016. In Vivo Deletion of the Cebpa +37 kb Enhancer Markedly Reduces Cebpa mRNA in Myeloid Progenitors but Not in Non-Hematopoietic Tissues to Impair Granulopoiesis. *PLoS One*. 11:e0150809.

Guo, H., O. Ma, N.A. Speck, and A.D. Friedman. 2012. Runx1 deletion or dominant inhibition reduces Cebpa transcription via conserved promoter and distal enhancer sites to favor monopoiesis over granulopoiesis. *Blood*. 119:4408-4418.

Guo, S.K., M.F. Shen, H.W. Yao, and Y.S. Liu. 2018. Enhanced Expression of TGFBI Promotes the Proliferation and Migration of Glioma Cells. *Cell Physiol Biochem*. 49:1097-1109.

Gurka, S., E. Hartung, M. Becker, and R.A. Kroczek. 2015. Mouse Conventional Dendritic Cells Can be Universally Classified Based on the Mutually Exclusive Expression of XCR1 and SIRPalpha. *Front Immunol*. 6:35.

Haakonsson, A.K., M. Stahl Madsen, R. Nielsen, A. Sandelin, and S. Mandrup. 2013. Acute genome-wide effects of rosiglitazone on PPARgamma transcriptional networks in adipocytes. *Mol Endocrinol*. 27:1536-1549.

Han, B., H. Cai, Y. Chen, B. Hu, H. Luo, Y. Wu, and J. Wu. 2015. The role of TGFBI (betaig-H3) in gastrointestinal tract tumorigenesis. *Mol Cancer*. 14:64.

Hannah, R., A. Joshi, N.K. Wilson, S. Kinston, and B. Gottgens. 2011. A compendium of genome-wide hematopoietic transcription factor maps supports the identification of gene regulatory control mechanisms. *Exp Hematol*. 39:531-541.

Hansen, K.D., Z. Wu, R.A. Irizarry, and J.T. Leek. 2011. Sequencing technology does not eliminate biological variability. *Nat Biotechnol*. 29:572-573.

Hansmann, L., A. Han, L. Penter, M. Liedtke, and M.M. Davis. 2017. Clonal Expansion and Interrelatedness of Distinct B-Lineage Compartments in Multiple Myeloma Bone Marrow. *Cancer Immunol Res*. 5:744-754.

Hanssen, L.L.P., M.T. Kassouf, A.M. Oudelaar, D. Biggs, C. Preece, D.J. Downes, M. Gosden, J.A. Sharpe,

J.A. Sloane-Stanley, J.R. Hughes, B. Davies, and D.R. Higgs. 2017. Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat Cell Biol*. 19:952-961.

Hasemann, M.S., F.K. Lauridsen, J. Waage, J.S. Jakobsen, A.K. Frank, M.B. Schuster, N. Rapin, F.O. Bagger, P.S. Hoppe, T. Schroeder, and B.T. Porse. 2014. C/EBPalpha is required for long-term self-renewal and lineage priming of hematopoietic stem cells and for the maintenance of epigenetic configurations in multipotent progenitors. *PLoS Genet*. 10:e1004079.

Hawkins, E.D., D. Duarte, O. Akinduro, R.A. Khorshed, D. Passaro, M. Nowicka, L. Straszkowski, M.K. Scott, S. Rothery, N. Ruivo, K. Foster, M. Waibel, R.W. Johnstone, S.J. Harrison, D.A. Westerman, H. Quach, J. Gribben, M.D. Robinson, L.E. Purton, D. Bonnet, and C. Lo Celso. 2016. T-cell acute leukaemia exhibits dynamic interactions with bone marrow microenvironments. *Nature*. 538:518-522.

Heath, H., C. Ribeiro de Almeida, F. Sleutels, G. Dingjan, S. van de Nobelen, I. Jonkers, K.W. Ling, J. Gribnau, R. Renkawitz, F. Grosveld, R.W. Hendriks, and N. Galjart. 2008. CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *EMBO J*. 27:2839-2850.

Heath, V., H.C. Suh, M. Holman, K. Renn, J.M. Gooya, S. Parkin, K.D. Klarmann, M. Ortiz, P. Johnson, and J. Keller. 2004. C/EBPalpha deficiency results in hyperproliferation of hematopoietic progenitor cells and disrupts macrophage development in vitro and in vivo. *Blood*. 104:1639-1647.

Heckl, D., M.S. Kowalczyk, D. Yudovich, R. Belizaire, R.V. Puram, M.E. McConkey, A. Thielke, J.C. Aster, A. Regev, and B.L. Ebert. 2014. Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat Biotechnol*. 32:941-946.

Heger, L., S. Balk, J.J. Luhr, G.F. Heidkamp, C.H.K. Lehmann, L. Hatscher, A. Purbojo, A. Hartmann, F. Garcia-Martin, S.I. Nishimura, R. Cesnjevar, F. Nimmerjahn, and D. Dudziak. 2018. CLEC10A Is a Specific Marker for Human CD1c(+) Dendritic Cells and Enhances Their Toll-Like Receptor 7/8-Induced Cytokine Secretion. *Front Immunol*. 9:744.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, and C.K. Glass. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 38:576-589.

Hildner, K., B.T. Edelson, W.E. Purtha, M. Diamond, H. Matsushita, M. Kohyama, B. Calderon, B.U. Schraml, E.R. Unanue, M.S. Diamond, R.D. Schreiber, T.L. Murphy, and K.M. Murphy. 2008. Batf3 deficiency reveals a critical role for CD8alpha+ dendritic cells in cytotoxic T cell immunity. *Science*. 322:1097-1100.

Hirai, H., P. Zhang, T. Dayaram, C.J. Hetherington, S. Mizuno, J. Imanishi, K. Akashi, and D.G. Tenen. 2006. C/EBPbeta is required for 'emergency' granulopoiesis. *Nat Immunol*. 7:732-739.

Hock, H., M.J. Hamblen, H.M. Rooke, D. Traver, R.T. Bronson, S. Cameron, and S.H. Orkin. 2003. Intrinsic requirement for zinc finger transcription factor Gfi-1 in neutrophil differentiation. *Immunity*. 18:109-120.

Hodonsky, C.J., D. Jain, U.M. Schick, J.V. Morrison, L. Brown, C.P. McHugh, C. Schurmann, D.D. Chen, Y.M. Liu, P.L. Auer, C.A. Laurie, K.D. Taylor, B.L. Browning, Y. Li, G. Papanicolaou, J.I. Rotter, R. Kurita, Y. Nakamura, S.R. Browning, R.J.F. Loos, K.E. North, C.C. Laurie, T.A. Thornton, N. Pankratz, D.E. Bauer, T. Sofer, and A.P. Reiner. 2017. Genome-wide association study of red blood cell traits in Hispanics/Latinos: The Hispanic Community Health Study/Study of Latinos. *PLoS Genet*. 13:e1006760.

Hong, S., A.M. Skaist, S.J. Wheelan, and A.D. Friedman. 2011. AP-1 protein induction during monopoiesis favors C/EBP: AP-1 heterodimers over C/EBP homodimerization and stimulates FosB transcription. *J Leukoc Biol*. 90:643-651.

Hoppe, P.S., M. Schwarzfischer, D. Loeffler, K.D. Kokkaliaris, O. Hilsenbeck, N. Moritz, M. Endele, A. Filipczyk, A. Gambardella, N. Ahmed, M. Etzrodt, D.L. Coutu, M.A. Rieger, C. Marr, M.K. Strasser, B. Schauberger, I. Burtscher, O. Ermakova, A. Burger, H. Lickert, C. Nerlov, F.J. Theis, and T. Schroeder. 2016. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature*. 535:299-302.

Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *J Educ Psychol*. 24:417-441.

Hu, J., O. Winqvist, A. Flores-Morales, A.C. Wikstrom, and G. Norstedt. 2009. SOCS2 influences LPS induced human monocyte-derived dendritic cell maturation. *PLoS One*. 4:e7178.

Infante, A., U. Laresgoiti, J. Fernandez-Rueda, A. Fullaondo, J. Galan, R. Diaz-Uriarte, M. Malumbres, S.J. Field, and A.M. Zubiaga. 2008. E2F2 represses cell cycle regulators to maintain quiescence. *Cell Cycle*. 7:3915-3927.

Ishikawa, F., H. Niiro, T. Iino, S. Yoshida, N. Saito, S. Onohara, T. Miyamoto, H. Minagawa, S. Fujii, L.D. Shultz, M. Harada, and K. Akashi. 2007. The developmental program of human dendritic cells is operated independently of conventional myeloid and lymphoid pathways. *Blood*. 110:3591-3660.

Ishikawa, J., Y. Takahashi, M. Hazawa, Y. Fukushi, A. Yoshizawa, and I. Kashiwakura. 2012. Suppressive effects of liquid crystal compounds on the growth of U937 human leukemic monocyte lymphoma cells. *Cancer Cell Int*. 12:3.

Jaitin, D.A., A. Weiner, I. Yofe, D. Lara-Astiaso, H. Keren-Shaul, E. David, T.M. Salame, A. Tanay, A. van Oudenaarden, and I. Amit. 2016. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell*. 167:1883-1896 e1815.

Jamieson, C.H., L.E. Ailles, S.J. Dylla, M. Muijtjens, C. Jones, J.L. Zehnder, J. Gotlib, K. Li, M.G. Manz, A. Keating, C.L. Sawyers, and I.L. Weissman. 2004. Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. *N Engl J Med*. 351:657-667.

Jang, S.K., H.G. Krausslich, M.J. Nicklin, G.M. Duke, A.C. Palmenberg, and E. Wimmer. 1988. A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol*. 62:2636-2643.

Jiapaer, Z., G. Li, D. Ye, M. Bai, J. Li, X. Guo, Y. Du, D. Su, W. Jia, W. Chen, G. Wang, Y. Yu, F. Zhu, X. Wan, and J. Kang. 2018. LincU Preserves Naive Pluripotency by Restricting ERK Activity in Embryonic Stem Cells. *Stem Cell Reports*. 11:395-409.

Johnson, D.S., A. Mortazavi, R.M. Myers, and B. Wold. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science*. 316:1497-1502.

Ju, X., M. Zenke, D.N. Hart, and G.J. Clark. 2008. CD300a/c regulate type I interferon and TNF-alpha secretion by human plasmacytoid dendritic cells stimulated with TLR7 and TLR9 ligands. *Blood*. 112:1184-1194.

Kameda, Y., M. Hanayama, A. Kishimoto, M. Kume, K. Yamamoto, and N. Matsumoto. 2016. Dendritic cell inhibitory receptor 4 (DCIR4) is preferentially expressed on inflammatory and patrolling monocytes. *Biochem Biophys Res Commun*. 480:215-221.

Kampen, K.R. 2012. The discovery and early understanding of leukemia. *Leuk Res*. 36:6-13.

Karamitros, D., B. Stoilova, Z. Aboukhalil, F. Hamey, A. Reinisch, M. Samitsch, L. Quek, G. Otto, E. Repapi, J. Doondeea, B. Usukhbayar, J. Calvo, S. Taylor, N. Goardon, E. Six, F. Pflumio, C. Porcher, R. Majeti, B. Gottgens, and P. Vyas. 2018. Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat Immunol*. 19:85-97.

Karsunky, H., M. Merad, A. Cozzio, I.L. Weissman, and M.G. Manz. 2003. Flt3 ligand regulates dendritic cell development from Flt3+ lymphoid and myeloid-committed progenitors to Flt3+ dendritic cells in vivo. *J Exp Med*. 198:305-313.

Karsunky, H., H. Zeng, T. Schmidt, B. Zevnik, R. Kluge, K.W. Schmid, U. Duhrsen, and T. Moroy. 2002. Inflammatory reactions and severe neutropenia in mice lacking the transcriptional repressor Gfi1. *Nat Genet*. 30:295-300.

Kato, N., J. Kitaura, N. Doki, Y. Komeno, N. Watanabe-Okochi, K. Togami, F. Nakahara, T. Oki, Y. Enomoto, Y. Fukuchi, H. Nakajima, Y. Harada, H. Harada, and T. Kitamura. 2011. Two types of C/EBPalpha mutations play distinct but collaborative roles in leukemogenesis: lessons from clinical data and BMT models. *Blood*. 117:221-233.

Katzerke, C., V. Madan, D. Gerloff, D. Brauer-Hartmann, J.U. Hartmann, A.A. Wurm, C. Muller-Tidow, S. Schnittger, D.G. Tenen, D. Niederwieser, and G. Behre. 2013. Transcription factor C/EBPalpha-induced microRNA-30c inactivates Notch1 during granulopoiesis and is downregulated in acute myeloid leukemia. *Blood*. 122:2433-2442.

Kharchenko, P.V., L. Silberstein, and D.T. Scadden. 2014. Bayesian approach to single-cell differential expression analysis. *Nat Methods*. 11:740-742.

Khoury, H.J., R.H. Collins, Jr., W. Blum, P.S. Stiff, L. Elias, J.S. Lebkowski, A. Reddy, K.P. Nishimoto, D. Sen, E.D. Wirth, 3rd, C.C. Case, and J.F. DiPersio. 2017. Immune responses and long-term disease recurrence status after telomerase-based dendritic cell immunotherapy in patients with acute myeloid leukemia. *Cancer*. 123:3061-3072.

Khoury, J.D. 2018. Blastic Plasmacytoid Dendritic Cell Neoplasm. *Curr Hematol Malig Rep*. 13:477-483.

Kiel, M.J., O.H. Yilmaz, T. Iwashita, O.H. Yilmaz, C. Terhorst, and S.J. Morrison. 2005. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell*. 121:1109-1121.

Kim, T.G., S. Kim, S. Jung, M. Kim, B. Yang, M.G. Lee, and H.P. Kim. 2017. CCCTC-binding factor is essential to the maintenance and quiescence of hematopoietic stem cells in mice. *Exp Mol Med*. 49:e371.

Kirstetter, P., M.B. Schuster, O. Bereshchenko, S. Moore, H. Dvinge, E. Kurz, K. Theilgaard-Monch, R. Mansson, T.A. Pedersen, T. Pabst, E. Schrock, B.T. Porse, S.E. Jacobsen, P. Bertone, D.G. Tenen, and C. Nerlov. 2008. Modeling of C/EBPalpha mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells. *Cancer Cell*. 13:299-310.

Klamer, S.E., Y.L. Dorland, M. Kleijer, D. Geerts, W.E. Lento, C.E. van der Schoot, M. von Lindern, and C. Voermans. 2018. TGFBI Expressed by Bone Marrow Niche Cells and Hematopoietic Stem and Progenitor Cells Regulates Hematopoiesis. *Stem Cells Dev*. 27:1494-1506.

Knapp, D., C.A. Hammond, F. Wang, N. Aghaeepour, P.H. Miller, P.A. Beer, D. Pellacani, M. VanInsberghe, C. Hansen, S.C. Bendall, G.P. Nolan, and C.J. Eaves. 2019. A topological view of human CD34(+) cell state trajectories from integrated single-cell output and proteomic data. *Blood*. 133:927-939.

Knoepfler, P.S., D.B. Sykes, M. Pasillas, and M.P. Kamps. 2001. HoxB8 requires its Pbx-interaction motif to block differentiation of primary myeloid progenitors and of most cell line models of myeloid differentiation. *Oncogene*. 20:5440-5448.

Kolb, H.J. 1998. Donor leukocyte transfusions for treatment of leukemic relapse after bone marrow transplantation. EBMT Immunology and Chronic Leukemia Working Parties. *Vox Sang*. 74 Suppl 2:321-329.

Kondo, M., I.L. Weissman, and K. Akashi. 1997. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell*. 91:661-672.

Koschmieder, S., F. D'Alo, H. Radomska, C. Schoneich, J.S. Chang, M. Konopleva, S. Kobayashi, E. Levantini, N. Suh, A. Di Ruscio, M.T. Voso, J.C. Watt, R. Santhanam, B. Sargin, H. Kantarjian, M. Andreeff, M.B. Sporn, D. Perrotti, W.E. Berdel, C. Muller-Tidow, H. Serve, and D.G. Tenen.

2007. CDDO induces granulocytic differentiation of myeloid leukemic blasts through translational up-regulation of p42 CCAAT enhancer binding protein alpha. *Blood*. 110:3695-3705.

Kowenz-Leutz, E., A. Schuetz, Q. Liu, M. Knoblich, U. Heinemann, and A. Leutz. 2016. Functional interaction of CCAAT/enhancer-binding-protein-alpha basic region mutants with E2F transcription factors and DNA. *Biochim Biophys Acta*. 1859:841-847.

Krivtsov, A.V., D. Twomey, Z. Feng, M.C. Stubbs, Y. Wang, J. Faber, J.E. Levine, J. Wang, W.C. Hahn, D.G. Gilliland, T.R. Golub, and S.A. Armstrong. 2006. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature*. 442:818-822.

Kuleshov, M.V., M.R. Jones, A.D. Rouillard, N.F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S.L. Jenkins, K.M. Jagodnik, A. Lachmann, M.G. McDermott, C.D. Monteiro, G.W. Gundersen, and A. Ma'ayan. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 44:W90-97.

Kummalue, T., and A.D. Friedman. 2003. Cross-talk between regulators of myeloid development: C/EBPalpha binds and activates the promoter of the PU.1 gene. *J Leukoc Biol*. 74:464-470.

Kuo, Y.H., S.K. Zaidi, S. Gornostaeva, T. Komori, G.S. Stein, and L.H. Castilla. 2009. Runx2 induces acute myeloid leukemia in cooperation with Cbfbeta-SMMHC in mice. *Blood*. 113:3323-3332.

Kuo, Y.Y., H.A. Hou, Y.K. Chen, L.Y. Li, P.H. Chen, M.H. Tseng, C.F. Huang, F.Y. Lee, M.C. Liu, C.W. Liu, W.C. Chou, C.Y. Liu, J.L. Tang, M. Yao, and H.F. Tien. 2014. The N-terminal CEBPA mutant in acute myeloid leukemia impairs CXCR4 expression. *Haematologica*. 99:1799-1807.

Kurotaki, D., W. Kawase, H. Sasaki, J. Nakabayashi, A. Nishiyama, H.C. Morse, 3rd, K. Ozato, Y. Suzuki, and T. Tamura. 2019. Epigenetic control of early dendritic cell lineage specification by the transcription factor IRF8 in mice. *Blood*. 133:1803-1813.

Kurotaki, D., J. Nakabayashi, A. Nishiyama, H. Sasaki, W. Kawase, N. Kaneko, K. Ochiai, K. Igarashi, K. Ozato, Y. Suzuki, and T. Tamura. 2018. Transcription Factor IRF8 Governs Enhancer Landscape Dynamics in Mononuclear Phagocyte Progenitors. *Cell Rep*. 22:2628-2641.

Kurotaki, D., N. Osato, A. Nishiyama, M. Yamamoto, T. Ban, H. Sato, J. Nakabayashi, M. Umehara, N. Miyake, N. Matsumoto, M. Nakazawa, K. Ozato, and T. Tamura. 2013. Essential role of the IRF8-KLF4 transcription factor cascade in murine monocyte differentiation. *Blood*. 121:1839-1849.

Kurotaki, D., M. Yamamoto, A. Nishiyama, K. Uno, T. Ban, M. Ichino, H. Sasaki, S. Matsunaga, M. Yoshinari, A. Ryo, M. Nakazawa, K. Ozato, and T. Tamura. 2014. IRF8 inhibits C/EBPalpha activity to restrain mononuclear phagocyte progenitors from differentiating into neutrophils. *Nat Commun*. 5:4978.

Laiosa, C.V., M. Stadtfeld, H. Xie, L. de Andres-Aguayo, and T. Graf. 2006. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. *Immunity*. 25:731-744.

Lapidot, T., C. Sirard, J. Vormoor, B. Murdoch, T. Hoang, J. Caceres-Cortes, M. Minden, B. Paterson, M.A. Caligiuri, and J.E. Dick. 1994. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*. 367:645-648.

Lauden, L., J. Siewiera, W. Boukouaci, K. Ramgolam, S. Mourah, C. Lebbe, D. Charron, F. Aoudjit, N. Jabrane-Ferrat, and R. Al-Daccak. 2014. TGF-beta-induced (TGFBI) protein in melanoma: a signature of high metastatic potential. *J Invest Dermatol*. 134:1675-1685.

Lavallee, V.P., J. Krosl, S. Lemieux, G. Boucher, P. Gendron, C. Pabst, I. Boivin, A. Marinier, C.J. Guidos, S. Meloche, J. Hebert, and G. Sauvageau. 2016. Chemo-genomic interrogation of CEBPA mutated AML reveals recurrent CSF3R mutations and subgroup sensitivity to JAK inhibitors. *Blood*. 127:3054-3061.

Le Mercier, I., D. Poujol, A. Sanlaville, V. Sisirak, M. Gobert, I. Durand, B. Dubois, I. Treilleux, J. Marvel, J. Vlach, J.Y. Blay, N. Bendriss-Vermare, C. Caux, I. Puisieux, and N. Goutagny. 2013. Tumor promotion by intratumoral plasmacytoid dendritic cells is reversed by TLR7 ligand treatment. *Cancer Res*. 73:4629-4640.

Lee, J., Y.J. Zhou, W. Ma, W. Zhang, A. Aljoufi, T. Luh, K. Lucero, D. Liang, M. Thomsen, G. Bhagat, Y. Shen, and K. Liu. 2017. Lineage specification of human dendritic cells is marked by IRF8 expression in hematopoietic stem cells and multipotent progenitors. *Nat Immunol*. 18:877-888.

Lewis, K.L., M.L. Caton, M. Bogunovic, M. Greter, L.T. Grajkowska, D. Ng, A. Klinakis, I.F. Charo, S. Jung, J.L. Gommerman, Ivanov, II, K. Liu, M. Merad, and B. Reizis. 2011. Notch2 receptor signaling controls functional differentiation of dendritic cells in the spleen and intestine. *Immunity*. 35:780-791.

Ley, T.J., L. Ding, M.J. Walter, M.D. McLellan, T. Lamprecht, D.E. Larson, C. Kandoth, J.E. Payton, J. Baty, J. Welch, C.C. Harris, C.F. Lichti, R.R. Townsend, R.S. Fulton, D.J. Dooling, D.C. Koboldt, H. Schmidt, Q. Zhang, J.R. Osborne, L. Lin, M. O'Laughlin, J.F. McMichael, K.D. Delehaunty, S.D. McGrath, L.A. Fulton, V.J. Magrini, T.L. Vickery, J. Hundal, L.L. Cook, J.J. Conyers, G.W. Swift, J.P. Reed, P.A. Alldredge, T. Wylie, J. Walker, J. Kalicki, M.A. Watson, S. Heath, W.D. Shannon, N. Varghese, R. Nagarajan, P. Westervelt, M.H. Tomasson, D.C. Link, T.A. Graubert, J.F. DiPersio, E.R. Mardis, and R.K. Wilson. 2010. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med*. 363:2424-2433.

Libura, M., M. Pawelczyk, I. Florek, K. Matiakowska, B. Jazwiec, K. Borg, I. Solarska, M. Zawada, S. Czekalska, J. Libura, Z. Salamanczuk, M. Jakobczyk, B. Mucha, E. Duszenko, K. Soszynska, K. Karabin, B. Piatkowska-Jakubas, M. Calbecka, J. Gajkowska-Kulig, G. Gadomska, M. Kielbinski, A. Ejduk, D. Kata, S. Grosicki, S. Kyrcz-Krzemien, K. Warzocha, K. Kuliczkowski, A. Skotnicki, W.W. Jerzejczak, and O. Haus. 2015. CEBPA copy number variations in normal karyotype acute myeloid leukemia: Possible role of breakpoint-associated microhomology and chromatin status in CEBPA mutagenesis. *Blood Cells Mol Dis*. 55:284-292.

Lichtenberg, J., E.F. Heuston, T. Mishra, C.A. Keller, R.C. Hardison, and D.M. Bodine. 2016. SBR-Blood: systems biology repository for hematopoietic cells. *Nucleic Acids Res*. 44:D925-931.

Lieu, Y.K., and E.P. Reddy. 2012. Impaired adult myeloid progenitor CMP and GMP cell function in conditional c-myb-knockout mice. *Cell Cycle*. 11:3504-3512.

Lin, Q., H. Chauvistre, I.G. Costa, E.G. Gusmao, S. Mitzka, S. Hanzelmann, B. Baying, T. Klisch, R. Moriggl, B. Hennuy, H. Smeets, K. Hoffmann, V. Benes, K. Sere, and M. Zenke. 2015. Epigenetic program and transcription factor circuitry of dendritic cell development. *Nucleic Acids Res*. 43:9680-9693.

Ling, G.S., J. Bennett, K.J. Woollard, M. Szajna, L. Fossati-Jimack, P.R. Taylor, D. Scott, G. Franzoso, H.T. Cook, and M. Botto. 2014. Integrin CD11b positively regulates TLR4-induced signalling pathways in dendritic cells but not in macrophages. *Nat Commun*. 5:3039.

Liu, F., H.Y. Wu, R. Wesselschmidt, T. Kornaga, and D.C. Link. 1996. Impaired production and increased apoptosis of neutrophils in granulocyte colony-stimulating factor receptor-deficient mice. *Immunity*. 5:491-501.

Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15:550.

Lu, R., N.F. Neff, S.R. Quake, and I.L. Weissman. 2011. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat Biotechnol*. 29:928-933.

Ma, O., S. Hong, H. Guo, G. Ghiaur, and A.D. Friedman. 2014. Granulopoiesis requires increased C/EBPalpha compared to monopoiesis, correlated with elevated Cebpa in immature G-CSF

receptor versus M-CSF receptor expressing cells. *PLoS One*. 9:e95784.

Macaulay, I.C., V. Svensson, C. Labalette, L. Ferreira, F. Hamey, T. Voet, S.A. Teichmann, and A. Cvejic. 2016. Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep*. 14:966-977.

Machado, R.A., H.S.B. Moreira, S.N. de Aquino, H. Martelli-Junior, S.R. de Almeida Reis, D.C. Persuhn, T. Wu, Y. Yuan, and R.D. Coletta. 2016. Interactions between RAD51 rs1801321 and maternal cigarette smoking as risk factor for nonsyndromic cleft lip with or without cleft palate. *Am J Med Genet A*. 170A:536-539.

Macosko, E.Z., A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A.R. Bialas, N. Kamitaki, E.M. Martersteck, J.J. Trombetta, D.A. Weitz, J.R. Sanes, A.K. Shalek, A. Regev, and S.A. McCarroll. 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*. 161:1202-1214.

Mardis, E.R., L. Ding, D.J. Dooling, D.E. Larson, M.D. McLellan, K. Chen, D.C. Koboldt, R.S. Fulton, K.D. Delehaunty, S.D. McGrath, L.A. Fulton, D.P. Locke, V.J. Magrini, R.M. Abbott, T.L. Vickery, J.S. Reed, J.S. Robinson, T. Wylie, S.M. Smith, L. Carmichael, J.M. Eldred, C.C. Harris, J. Walker, J.B. Peck, F. Du, A.F. Dukes, G.E. Sanderson, A.M. Brummett, E. Clark, J.F. McMichael, R.J. Meyer, J.K. Schindler, C.S. Pohl, J.W. Wallis, X. Shi, L. Lin, H. Schmidt, Y. Tang, C. Haipek, M.E. Wiechert, J.V. Ivy, J. Kalicki, G. Elliott, R.E. Ries, J.E. Payton, P. Westervelt, M.H. Tomasson, M.A. Watson, J. Baty, S. Heath, W.D. Shannon, R. Nagarajan, D.C. Link, M.J. Walter, T.A. Graubert, J.F. DiPersio, R.K. Wilson, and T.J. Ley. 2009. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med*. 361:1058-1066.

Mavin, E., L. Nicholson, S. Rafez Ahmed, F. Gao, A. Dickinson, and X.N. Wang. 2017. Human Regulatory T Cells Mediate Transcriptional Modulation of Dendritic Cell Function. *J Immunol*. 198:138-146.

Maxson, J.E., R.E. Ries, Y.C. Wang, R.B. Gerbing, E.A. Kolb, S.L. Thompson, J.M. Guidry Auvil, M.A. Marra, Y. Ma, Z. Zong, A.J. Mungall, R. Moore, W. Long, P. Gesuwan, T.M. Davidsen, L.C. Hermida, S.B. Hughes, J.E. Farrar, J.P. Radich, M.A. Smith, D.S. Gerhard, A.S. Gamis, T.A. Alonzo, and S. Meshinchi. 2016. CSF3R mutations have a high degree of overlap with CEBPA mutations in pediatric AML. *Blood*. 127:3094-3098.

McInnes, L., J. Healy, and J. Melville. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *https://arxiv.org/abs/1802.03426*.

Miao, Z., K. Deng, X. Wang, and X. Zhang. 2018. DEsingle for detecting three types of differential expression in single-cell RNA-seq data. *Bioinformatics*. 34:3223-3224.

Miller, J.C., B.D. Brown, T. Shay, E.L. Gautier, V. Jojic, A. Cohain, G. Pandey, M. Leboeuf, K.G. Elpek, J. Helft, D. Hashimoto, A. Chow, J. Price, M. Greter, M. Bogunovic, A. Bellemare-Pelletier, P.S. Frenette, G.J. Randolph, S.J. Turley, M. Merad, and C. Immunological Genome. 2012. Deciphering the transcriptional network of the dendritic cell lineage. *Nat Immunol*. 13:888-899.

Miller, M., J.D. Shuman, T. Sebastian, Z. Dauter, and P.F. Johnson. 2003. Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. *J Biol Chem*. 278:15178-15184.

Mittelbrunn, M., G. Martinez del Hoyo, M. Lopez-Bravo, N.B. Martin-Cofreces, A. Scholer, S. Hugues, L. Fetler, S. Amigorena, C. Ardavin, and F. Sanchez-Madrid. 2009. Imaging of plasmacytoid dendritic cell interactions with T cells. *Blood*. 113:75-84.

Miyamoto, T., I.L. Weissman, and K. Akashi. 2000. AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. *Proc Natl Acad Sci U S A*. 97:7521-7526.

Mizuguchi, H., Z. Xu, A. Ishii-Watabe, E. Uchida, and T. Hayakawa. 2000. IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. *Mol Ther*. 1:376-382.

Mo, X.D., Y. Wang, X.H. Zhang, L.P. Xu, C.H. Yan, H. Chen, Y.H. Chen, Y.Z. Qin, K.Y. Liu, and X.J. Huang. 2018. Interferon-alpha Is Effective for Treatment of Minimal Residual Disease in Patients with t(8;21) Acute Myeloid Leukemia After Allogeneic Hematopoietic Stem Cell Transplantation: Results of a Prospective Registry Study. *Oncologist*. 23:1349-1357.

Moignard, V., and B. Gottgens. 2014. Transcriptional mechanisms of cell fate decisions revealed by single cell expression profiling. *Bioessays*. 36:419-426.

Moignard, V., I.C. Macaulay, G. Swiers, F. Buettner, J. Schutte, F.J. Calero-Nieto, S. Kinston, A. Joshi, R. Hannah, F.J. Theis, S.E. Jacobsen, M.F. de Bruijn, and B. Gottgens. 2013. Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat Cell Biol*. 15:363-372.

Moignard, V., S. Woodhouse, J. Fisher, and B. Gottgens. 2013. Transcriptional hierarchies regulating early blood cell development. *Blood Cells Mol Dis*. 51:239-247.

Moran-Crusio, K., L. Reavie, A. Shih, O. Abdel-Wahab, D. Ndiaye-Lobry, C. Lobry, M.E. Figueroa, A. Vasanthakumar, J. Patel, X. Zhao, F. Perna, S. Pandey, J. Madzo, C. Song, Q. Dai, C. He, S. Ibrahim, M. Beran, J. Zavadil, S.D. Nimer, A. Melnick, L.A. Godley, I. Aifantis, and R.L. Levine. 2011. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*. 20:11-24.

Mucenski, M.L., K. McLain, A.B. Kier, S.H. Swerdlow, C.M. Schreiner, T.A. Miller, D.W. Pietryga, W.J. Scott, Jr., and S.S. Potter. 1991. A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. *Cell*. 65:677-689.

Murphy, T.L., G.E. Grajales-Reyes, X. Wu, R. Tussiwand, C.G. Briseno, A. Iwata, N.M. Kretzer, V. Durai, and K.M. Murphy. 2016. Transcriptional Control of Dendritic Cell Development. *Annu Rev Immunol*. 34:93-119.

Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*. 320:1344-1349.

Naik, S.H., L. Perie, E. Swart, C. Gerlach, N. van Rooij, R.J. de Boer, and T.N. Schumacher. 2013. Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature*. 496:229-232.

Narita, M., M. Takahashi, A. Liu, K. Nikkuni, T. Furukawa, K. Toba, S. Koyama, K. Takai, M. Sanada, and Y. Aizawa. 2001. Leukemia blast-induced T-cell anergy demonstrated by leukemia-derived dendritic cells in acute myelogenous leukemia. *Exp Hematol*. 29:709-719.

Nerlov, C. 2004. C/EBPalpha mutations in acute myeloid leukaemias. *Nat Rev Cancer*. 4:394-400.

Nerlov, C. 2007. The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control. *Trends Cell Biol*. 17:318-324.

Nimmo, R.A., G.E. May, and T. Enver. 2015. Primed and ready: understanding lineage commitment through single cell analysis. *Trends Cell Biol*. 25:459-467.

Notta, F., S. Zandi, N. Takayama, S. Dobson, O.I. Gan, G. Wilson, K.B. Kaufmann, J. McLeod, E. Laurenti, C.F. Dunant, J.D. McPherson, L.D. Stein, Y. Dror, and J.E. Dick. 2016. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*. 351:aab2116.

Noubade, R., S. Majri-Morrison, and K.V. Tarbell. 2019. Beyond cDC1: Emerging Roles of DC Crosstalk in Cancer Immunity. *Front Immunol*. 10:1014.

Oguro, H., L. Ding, and S.J. Morrison. 2013. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell*.

13:102-116.

Oh, J., N. Wu, G. Baravalle, B. Cohn, J. Ma, B. Lo, I. Mellman, S. Ishido, M. Anderson, and J.S. Shin. 2013. MARCH1-mediated MHCII ubiquitination promotes dendritic cell selection of natural regulatory T cells. *J Exp Med*. 210:1069-1077.

Ohlsson, E., M.S. Hasemann, A. Willer, F.K. Lauridsen, N. Rapin, J. Jendholm, and B.T. Porse. 2014. Initiation of MLL-rearranged AML is dependent on C/EBPalpha. *J Exp Med*. 211:5-13.

Ohlsson, E., M.B. Schuster, M. Hasemann, and B.T. Porse. 2016. The multifaceted functions of C/EBPalpha in normal and malignant haematopoiesis. *Leukemia*. 30:767-775.

Onai, N., M.G. Manz, and M.A. Schmid. 2010. Isolation of common dendritic cell progenitors (CDP) from mouse bone marrow. *Methods Mol Biol*. 595:195-203.

Onai, N., A. Obata-Onai, M.A. Schmid, T. Ohteki, D. Jarrossay, and M.G. Manz. 2007. Identification of clonogenic common Flt3+M-CSFR+ plasmacytoid and conventional dendritic cell progenitors in mouse bone marrow. *Nat Immunol*. 8:1207-1216.

Onai, N., A. Obata-Onai, R. Tussiwand, A. Lanzavecchia, and M.G. Manz. 2006. Activation of the Flt3 signal transduction cascade rescues and enhances type I interferon-producing and dendritic cell development. *J Exp Med*. 203:227-238.

Onodera, K., T. Fujiwara, Y. Onishi, A. Itoh-Nakadai, Y. Okitsu, N. Fukuhara, K. Ishizawa, R. Shimizu, M. Yamamoto, and H. Harigae. 2016. GATA2 regulates dendritic cell differentiation. *Blood*. 128:508-518.

Opavsky, R., S.Y. Tsai, M. Guimond, A. Arora, J. Opavska, B. Becknell, M. Kaufmann, N.A. Walton, J.A. Stephens, S.A. Fernandez, N. Muthusamy, D.W. Felsher, P. Porcu, M.A. Caligiuri, and G. Leone. 2007. Specific tumor suppressor function for E2F2 in Myc-induced T cell lymphomagenesis. *Proc Natl Acad Sci U S A*. 104:15400-15405.

Osawa, M., K. Hanada, H. Hamada, and H. Nakauchi. 1996. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science*. 273:242-245.

Ottersbach, K., A. Smith, A. Wood, and B. Gottgens. 2010. Ontogeny of haematopoiesis: recent advances and open questions. *Br J Haematol*. 148:343-355.

Ouboussad, L., S. Kreuz, and P.F. Lefevre. 2013. CTCF depletion alters chromatin structure and transcription of myeloid-specific factors. *J Mol Cell Biol*. 5:308-322.

Ozawa, D., T. Yokobori, M. Sohda, M. Sakai, K. Hara, H. Honjo, H. Kato, T. Miyazaki, and H. Kuwano. 2016. TGFBI Expression in Cancer Stromal Cells is Associated with Poor Prognosis and Hematogenous Recurrence in Esophageal Squamous Cell Carcinoma. *Ann Surg Oncol*. 23:282-289.

Pabst, T., M. Eyholzer, J. Fos, and B.U. Mueller. 2009. Heterogeneity within AML with CEBPA mutations; only CEBPA double mutations, but not single CEBPA mutations are associated with favourable prognosis. *Br J Cancer*. 100:1343-1346.

Pabst, T., M. Eyholzer, S. Haefliger, J. Schardt, and B.U. Mueller. 2008. Somatic CEBPA mutations are a frequent second event in families with germline CEBPA mutations and familial acute myeloid leukemia. *J Clin Oncol*. 26:5088-5093.

Pabst, T., and B.U. Mueller. 2009. Complexity of CEBPA dysregulation in human acute myeloid leukemia. *Clin Cancer Res*. 15:5303-5307.

Pabst, T., B.U. Mueller, P. Zhang, H.S. Radomska, S. Narravula, S. Schnittger, G. Behre, W. Hiddemann, and D.G. Tenen. 2001. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. *Nat Genet*. 27:263-270.

Pan, Y.B., C.H. Zhang, S.Q. Wang, P.H. Ai, K. Chen, L. Zhu, Z.L. Sun, and D.F. Feng. 2018. Transforming growth factor beta induced (TGFBI) is a potential signature gene for mesenchymal subtype

high-grade glioma. *J Neurooncol*. 137:395-407.

Pandey, G., A. Cohain, J. Miller, and M. Merad. 2013. Decoding dendritic cell function through module and network analysis. *J Immunol Methods*. 387:71-80.

Papaemmanuil, E., M. Gerstung, L. Bullinger, V.I. Gaidzik, P. Paschka, N.D. Roberts, N.E. Potter, M. Heuser, F. Thol, N. Bolli, G. Gundem, P. Van Loo, I. Martincorena, P. Ganly, L. Mudie, S. McLaren, S. O'Meara, K. Raine, D.R. Jones, J.W. Teague, A.P. Butler, M.F. Greaves, A. Ganser, K. Dohner, R.F. Schlenk, H. Dohner, and P.J. Campbell. 2016. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med*. 374:2209-2221.

Park, S.J., T. Nakagawa, H. Kitamura, T. Atsumi, H. Kamon, S. Sawa, D. Kamimura, N. Ueda, Y. Iwakura, K. Ishihara, M. Murakami, and T. Hirano. 2004. IL-6 regulates in vivo dendritic cell differentiation through STAT3 activation. *J Immunol*. 173:3844-3854.

Paul, F., Y. Arkin, A. Giladi, D.A. Jaitin, E. Kenigsberg, H. Keren-Shaul, D. Winter, D. Lara-Astiaso, M. Gury, A. Weiner, E. David, N. Cohen, F.K. Lauridsen, S. Haas, A. Schlitzer, A. Mildner, F. Ginhoux, S. Jung, A. Trumpp, B.T. Porse, A. Tanay, and I. Amit. 2015. Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell*. 163:1663-1677.

Picelli, S., A.K. Bjorklund, O.R. Faridani, S. Sagasser, G. Winberg, and R. Sandberg. 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods*. 10:1096-1098.

Picelli, S., O.R. Faridani, A.K. Bjorklund, G. Winberg, S. Sagasser, and R. Sandberg. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc*. 9:171-181.

Pietras, E.M., D. Reynaud, Y.A. Kang, D. Carlin, F.J. Calero-Nieto, A.D. Leavitt, J.M. Stuart, B. Gottgens, and E. Passegue. 2015. Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell*. 17:35-46.

Pimanda, J.E., and B. Gottgens. 2010. Gene regulatory networks governing haematopoietic stem cell development and identity. *Int J Dev Biol*. 54:1201-1211.

Pimanda, J.E., K. Ottersbach, K. Knezevic, S. Kinston, W.Y. Chan, N.K. Wilson, J.R. Landry, A.D. Wood, A. Kolb-Kokocinski, A.R. Green, D. Tannahill, G. Lacaud, V. Kouskoff, and B. Gottgens. 2007. Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proc Natl Acad Sci U S A*. 104:17692-17697.

Pineault, N., C.D. Helgason, H.J. Lawrence, and R.K. Humphries. 2002. Differential expression of Hox, Meis1, and Pbx1 genes in primitive cells throughout murine hematopoietic ontogeny. *Exp Hematol*. 30:49-57.

Pinto, J.P., R.S.R. Machado, R. Magno, D.V. Oliveira, S. Machado, R.P. Andrade, J. Braganca, I. Duarte, and M.E. Futschik. 2018. StemMapper: a curated gene expression database for stem cell lineage analysis. *Nucleic Acids Res*. 46:D788-D793.

Pogenberg, V., L. Consani Textor, L. Vanhille, S.J. Holton, M.H. Sieweke, and M. Wilmanns. 2014. Design of a bZip transcription factor with homo/heterodimer-induced DNA-binding preference. *Structure*. 22:466-477.

Porse, B.T., T.A. Pedersen, X. Xu, B. Lindberg, U.M. Wewer, L. Friis-Hansen, and C. Nerlov. 2001. E2F repression by C/EBPalpha is required for adipogenesis and granulopoiesis in vivo. *Cell*. 107:247-258.

Preudhomme, C., C. Sagot, N. Boissel, J.M. Cayuela, I. Tigaud, S. de Botton, X. Thomas, E. Raffoux, C. Lamandin, S. Castaigne, P. Fenaux, H. Dombret, and A. Group. 2002. Favorable prognostic significance of CEBPA mutations in patients with de novo acute myeloid leukemia: a study from the Acute Leukemia French Association (ALFA). *Blood*. 100:2717-2723.

Pronk, C.J., D.J. Rossi, R. Mansson, J.L. Attema, G.L. Norddahl, C.K. Chan, M. Sigvardsson, I.L. Weissman, and D. Bryder. 2007. Elucidation of the phenotypic, functional, and molecular topography of a

myeloerythroid progenitor cell hierarchy. *Cell Stem Cell*. 1:428-442.

Psaila, B., N. Barkas, D. Iskander, A. Roy, S. Anderson, N. Ashley, V.S. Caputo, J. Lichtenberg, S. Loaiza, D.M. Bodine, A. Karadimitris, A.J. Mead, and I. Roberts. 2016. Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. *Genome Biol*. 17:83.

Pundhir, S., F.K. Bratt Lauridsen, M.B. Schuster, J.S. Jakobsen, Y. Ge, E.M. Schoof, N. Rapin, J. Waage, M.S. Hasemann, and B.T. Porse. 2018. Enhancer and Transcription Factor Dynamics during Myeloid Differentiation Reveal an Early Differentiation Block in Cebpa null Progenitors. *Cell Rep*. 23:2744-2757.

Pyzer, A.R., D.E. Avigan, and J. Rosenblatt. 2014. Clinical trials of dendritic cell-based cancer vaccines in hematologic malignancies. *Hum Vaccin Immunother*. 10:3125-3131.

Qi, X., J. Hong, L. Chaves, Y. Zhuang, Y. Chen, D. Wang, J. Chabon, B. Graham, K. Ohmori, Y. Li, and H. Huang. 2013. Antagonistic regulation by the transcription factors C/EBPalpha and MITF specifies basophil and mast cell fates. *Immunity*. 39:97-110.

Quintana, E., M. Shackleton, M.S. Sabel, D.R. Fullen, T.M. Johnson, and S.J. Morrison. 2008. Efficient tumour formation by single human melanoma cells. *Nature*. 456:593-598.

Quintana-Bustamente, O., S.L. Smith, E. Griessinger, Y. Reyal, J. Vargaftig, T.A. Lister, J. Fitzgibbon, and D. Bonnet. 2012. Overexpression of wild-type or mutant forms of *CEBPA* alter normal human hematopoiesis. *Leukemia*. 26(7):1537-1546.

Rangatia, J., R.K. Vangala, S.M. Singh, A.A. Peer Zada, A. Elsasser, A. Kohlmann, T. Haferlach, D.G. Tenen, W. Hiddemann, and G. Behre. 2003. Elevated c-Jun expression in acute myeloid leukemias inhibits C/EBPalpha DNA binding via leucine zipper domain interaction. *Oncogene*. 22:4760-4764.

Rangatia, J., R.K. Vangala, N. Treiber, P. Zhang, H. Radomska, D.G. Tenen, W. Hiddemann, and G. Behre. 2002. Downregulation of c-Jun expression by transcription factor C/EBPalpha is critical for granulocytic lineage commitment. *Mol Cell Biol*. 22:8681-8694.

Raser, J.M., and E.K. O'Shea. 2004. Control of stochasticity in eukaryotic gene expression. *Science*. 304:1811-1814.

Rathinam, C., R. Geffers, R. Yucel, J. Buer, K. Welte, T. Moroy, and C. Klein. 2005. The transcriptional repressor Gfi1 controls STAT3-dependent dendritic cell development and function. *Immunity*. 22:717-728.

Reckzeh, K., and J. Cammenga. 2010. Molecular mechanisms underlying deregulation of C/EBPalpha in acute myeloid leukemia. *Int J Hematol*. 91:557-568.

Reddy, V.A., A. Iwama, G. Iotzova, M. Schulz, A. Elsasser, R.K. Vangala, D.G. Tenen, W. Hiddemann, and G. Behre. 2002. Granulocyte inducer C/EBPalpha inactivates the myeloid master regulator PU.1: possible role in lineage commitment decisions. *Blood*. 100:483-490.

Redecke, V., R. Wu, J. Zhou, D. Finkelstein, V. Chaturvedi, A.A. High, and H. Hacker. 2013. Hematopoietic progenitor cell lines with myeloid and lymphoid potential. *Nat Methods*. 10:795-803.

Reizis, B., A. Bunin, H.S. Ghosh, K.L. Lewis, and V. Sisirak. 2011. Plasmacytoid dendritic cells: recent progress and open questions. *Annu Rev Immunol*. 29:163-183.

Renneville, A., N. Boissel, N. Gachard, D. Naguib, C. Bastard, S. de Botton, O. Nibourel, C. Pautas, O. Reman, X. Thomas, C. Gardin, C. Terre, S. Castaigne, C. Preudhomme, and H. Dombret. 2009. The favorable impact of CEBPA mutations in patients with acute myeloid leukemia is only observed in the absence of associated cytogenetic abnormalities and FLT3 internal duplication. *Blood*. 113:5090-5093.

Riddell, J., R. Gazit, B.S. Garrison, G. Guo, A. Saadatpour, P.K. Mandal, W. Ebina, P. Volchkov, G.C. Yuan, S.H. Orkin, and D.J. Rossi. 2014. Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. *Cell*. 157:549-564.

Ripperger, T., G. Manukjan, J. Meyer, S. Wolter, A. Schambach, J. Bohne, U. Modlich, Z. Li, B. Skawran, B. Schlegelberger, and D. Steinemann. 2015. The heteromeric transcription factor GABP activates the ITGAM/CD11b promoter and induces myeloid differentiation. *Biochim Biophys Acta*. 1849:1145-1154.

Robbins, S.H., T. Walzer, D. Dembele, C. Thibault, A. Defays, G. Bessou, H. Xu, E. Vivier, M. Sellars, P. Pierre, F.R. Sharp, S. Chan, P. Kastner, and M. Dalod. 2008. Novel insights into the relationships between dendritic cell subsets in human and mouse revealed by genome-wide expression profiling. *Genome Biol*. 9:R17.

Rocha-Perugini, V., G. Martinez Del Hoyo, J.M. Gonzalez-Granado, M. Ramirez-Huesca, V. Zorita, E. Rubinstein, C. Boucheix, and F. Sanchez-Madrid. 2017. CD9 Regulates Major Histocompatibility Complex Class II Trafficking in Monocyte-Derived Dendritic Cells. *Mol Cell Biol*. 37.

Rodrigues, P.F., L. Alberti-Servera, A. Eremin, G.E. Grajales-Reyes, R. Ivanek, and R. Tussiwand. 2018. Distinct progenitor lineages contribute to the heterogeneity of plasmacytoid dendritic cells. *Nat Immunol*. 19:711-722.

Roes, J., B.K. Choi, D. Power, P. Xu, and A.W. Segal. 2003. Granulocyte function in grancalcin-deficient mice. *Mol Cell Biol*. 23:826-830.

Rosa, F.F., C.F. Pires, I. Kurochkin, A.G. Ferreira, A.M. Gomes, L.G. Palma, K. Shaiv, L. Solanas, C. Azenha, D. Papatsenko, O. Schulz, C. Reis e Sousa, and C.F. Pereira. 2018. Direct reprogramming of fibroblasts into antigen-presenting dendritic cells. *Sci Immunol*. 3.

Rose, S. 2017. First-Ever CAR T-cell Therapy Approved in U.S. *Cancer Discov*. 7:OF1.

Rosenbauer, F., K. Wagner, J.L. Kutok, H. Iwasaki, M.M. Le Beau, Y. Okuno, K. Akashi, S. Fiering, and D.G. Tenen. 2004. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat Genet*. 36:624-630.

Ross, S.E., H.S. Radomska, B. Wu, P. Zhang, J.N. Winnay, L. Bajnok, W.S. Wright, F. Schaufele, D.G. Tenen, and O.A. MacDougald. 2004. Phosphorylation of C/EBPalpha inhibits granulopoiesis. *Mol Cell Biol*. 24:675-686.

Rothenberg-Thurley, M., S. Amler, D. Goerlich, T. Kohnke, N.P. Konstandin, S. Schneider, M.C. Sauerland, T. Herold, M. Hubmann, B. Ksienzyk, E. Zellmeier, S.K. Bohlander, M. Subklewe, A. Faldum, W. Hiddemann, J. Braess, K. Spiekermann, and K.H. Metzeler. 2018. Persistence of pre-leukemic clones during first remission and risk of relapse in acute myeloid leukemia. *Leukemia*. 32:1598-1608.

Roug, A.S., M.C. Hansen, L. Nederby, and P. Hokland. 2014. Diagnosing and following adult patients with acute myeloid leukaemia in the genomic age. *Br J Haematol*. 167:162-176.

Ruau, D., F.S. Ng, N.K. Wilson, R. Hannah, E. Diamanti, P. Lombard, S. Woodhouse, and B. Gottgens. 2013. Building an ENCODE-style data compendium on a shoestring. *Nat Methods*. 10:926.

Sanchez-Castillo, M., D. Ruau, A.C. Wilkinson, F.S. Ng, R. Hannah, E. Diamanti, P. Lombard, N.K. Wilson, and B. Gottgens. 2015. CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. *Nucleic Acids Res*. 43:D1117-1123.

Sanjuan-Pla, A., I.C. Macaulay, C.T. Jensen, P.S. Woll, T.C. Luis, A. Mead, S. Moore, C. Carella, S. Matsuoka, T. Bouriez Jones, O. Chowdhury, L. Stenson, M. Lutteropp, J.C. Green, R. Facchini, H. Boukarabila, A. Grover, A. Gambardella, S. Thongjuea, J. Carrelha, P. Tarrant, D. Atkinson, S.A. Clark, C. Nerlov, and S.E. Jacobsen. 2013. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*. 502:232-236.

Sardina, J.L., B. Di Stefano, and T. Graf. 2016. How does C/EBPalpha speed up cell reprogramming? *Cell*

Cycle. 15:2381-2382.

Sasaki, M., C.B. Knobbe, J.C. Munger, E.F. Lind, D. Brenner, A. Brustle, I.S. Harris, R. Holmes, A. Wakeham, J. Haight, A. You-Ten, W.Y. Li, S. Schalm, S.M. Su, C. Virtanen, G. Reifenberger, P.S. Ohashi, D.L. Barber, M.E. Figueroa, A. Melnick, J.C. Zuniga-Pflucker, and T.W. Mak. 2012. IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. *Nature*. 488:656-659.

Satpathy, A.T., C.G. Briseno, X. Cai, D.G. Michael, C. Chou, S. Hsiung, D. Bhattacharya, N.A. Speck, and T. Egawa. 2014. Runx1 and Cbfbeta regulate the development of Flt3+ dendritic cell progenitors and restrict myeloproliferative disorder. *Blood*. 123:2968-2977.

Sawai, C.M., V. Sisirak, H.S. Ghosh, E.Z. Hou, M. Ceribelli, L.M. Staudt, and B. Reizis. 2013. Transcription factor Runx2 controls the development and migration of plasmacytoid dendritic cells. *J Exp Med*. 210:2151-2159.

Schepers, H., A.T. Wierenga, D. van Gosliga, B.J. Eggen, E. Vellenga, and J.J. Schuringa. 2007. Reintroduction of C/EBPalpha in leukemic CD34+ stem/progenitor cells impairs self-renewal and partially restores myelopoiesis. *Blood*. 110:1317-1325.

Schinnerling, K., P. Garcia-Gonzalez, and J.C. Aguillon. 2015. Gene Expression Profiling of Human Monocyte-derived Dendritic Cells - Searching for Molecular Regulators of Tolerogenicity. *Front Immunol*. 6:528.

Schonheit, J., C. Kuhl, M.L. Gebhardt, F.F. Klett, P. Riemke, M. Scheller, G. Huang, R. Naumann, A. Leutz, C. Stocking, J. Priller, M.A. Andrade-Navarro, and F. Rosenbauer. 2013. PU.1 level-directed chromatin structure remodeling at the Irf8 gene drives dendritic cell commitment. *Cell Rep*. 3:1617-1628.

Schuermann, M., J.B. Hunter, G. Hennig, and R. Muller. 1991. Non-leucine residues in the leucine repeats of Fos and Jun contribute to the stability and determine the specificity of dimerization. *Nucleic Acids Res*. 19:739-746.

Schuettpelz, L.G., P.K. Gopalan, F.O. Giuste, M.P. Romine, R. van Os, and D.C. Link. 2012. Kruppel-like factor 7 overexpression suppresses hematopoietic stem and progenitor cell function. *Blood*. 120:2981-2989.

Schulte, R., N.K. Wilson, J.C. Prick, C. Cossetti, M.K. Maj, B. Gottgens, and D.G. Kent. 2015. Index sorting resolves heterogeneous murine hematopoietic stem cell populations. *Exp Hematol*. 43:803-811.

Schurch, C.M., C. Riether, and A.F. Ochsenbein. 2013. Dendritic cell-based immunotherapy for myeloid leukemias. *Front Immunol*. 4:496.

Schutte, J., H. Wang, S. Antoniou, A. Jarratt, N.K. Wilson, J. Riepsaame, F.J. Calero-Nieto, V. Moignard, S. Basilico, S.J. Kinston, R.L. Hannah, M.C. Chan, S.T. Nurnberg, W.H. Ouwehand, N. Bonzanni, M.F. de Bruijn, and B. Gottgens. 2016. An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability. *Elife*. 5:e11469.

Scialdone, A., Y. Tanaka, W. Jawaid, V. Moignard, N.K. Wilson, I.C. Macaulay, J.C. Marioni, and B. Gottgens. 2016. Resolving early mesoderm diversification through single-cell expression profiling. *Nature*. 535:289-293.

Scutera, S., T. Fraone, T. Musso, P. Cappello, S. Rossi, D. Pierobon, Z. Orinska, R. Paus, S. Bulfone-Paus, and M. Giovarelli. 2009. Survival and migration of human dendritic cells are regulated by an IFN-alpha-inducible Axl/Gas6 pathway. *J Immunol*. 183:3004-3013.

Seipel, K., M.T. Marques, M.A. Bozzini, C. Meinken, B.U. Mueller, and T. Pabst. 2016. Inactivation of the p53-KLF4-CEBPA Axis in Acute Myeloid Leukemia. *Clin Cancer Res*. 22:746-756.

Seita, J., D. Sahoo, D.J. Rossi, D. Bhattacharya, T. Serwold, M.A. Inlay, L.I. Ehrlich, J.W. Fathman, D.L. Dill, and I.L. Weissman. 2012. Gene Expression Commons: an open platform for absolute gene

expression profiling. *PLoS One*. 7:e40321.

Severe, N., N.M. Karabacak, K. Gustafsson, N. Baryawno, G. Courties, Y. Kfoury, K.D. Kokkaliaris, C. Rhee, D. Lee, E.W. Scadden, J.E. Garcia-Robledo, T. Brouse, M. Nahrendorf, M. Toner, and D.T. Scadden. 2019. Stress-Induced Changes in Bone Marrow Stromal Cell Populations Revealed through Single-Cell Protein Expression Mapping. *Cell Stem Cell*.

Sharifi-Zarchi, A., D. Gerovska, K. Adachi, M. Totonchi, H. Pezeshk, R.J. Taft, H.R. Scholer, H. Chitsaz, M. Sadeghi, H. Baharvand, and M.J. Arauzo-Bravo. 2017. DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. *BMC Genomics*. 18:964.

Shi, X., B.L. He, A.C. Ma, Y. Guo, Y. Chi, C.H. Man, W. Zhang, Y. Zhang, Z. Wen, T. Cheng, and A.Y. Leung. 2015. Functions of idh1 and its mutation in the regulation of developmental hematopoiesis in zebrafish. *Blood*. 125:2974-2984.

Shigematsu, H., B. Reizis, H. Iwasaki, S. Mizuno, D. Hu, D. Traver, P. Leder, N. Sakaguchi, and K. Akashi. 2004. Plasmacytoid dendritic cells activate lymphoid-specific genetic programs irrespective of their cellular origin. *Immunity*. 21:43-53.

Shih, L.Y., D.C. Liang, C.F. Huang, J.H. Wu, T.L. Lin, P.N. Wang, P. Dunn, M.C. Kuo, and T.C. Tang. 2006. AML patients with CEBPalpha mutations mostly retain identical mutant patterns but frequently change in allelic distribution at relapse: a comparative analysis on paired diagnosis and relapse samples. *Leukemia*. 20:604-609.

Shimizu, R., T. Kuroha, O. Ohneda, X. Pan, K. Ohneda, S. Takahashi, S. Philipsen, and M. Yamamoto. 2004. Leukemogenesis caused by incapacitated GATA-1 function. *Mol Cell Biol*. 24:10814-10825.

Sichien, D., C.L. Scott, L. Martens, M. Vanderkerken, S. Van Gassen, M. Plantinga, T. Joeris, S. De Prijck, L. Vanhoutte, M. Vanheerswynghels, G. Van Isterdael, W. Toussaint, F.B. Madeira, K. Vergote, W.W. Agace, B.E. Clausen, H. Hammad, M. Dalod, Y. Saeys, B.N. Lambrecht, and M. Guilliams. 2016. IRF8 Transcription Factor Controls Survival and Function of Terminally Differentiated Conventional and Plasmacytoid Dendritic Cells, Respectively. *Immunity*. 45:626-640.

Simonis, M., P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, B. van Steensel, and W. de Laat. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*. 38:1348-1354.

Sive, J.I., S. Basilico, R. Hannah, S.J. Kinston, F.J. Calero-Nieto, and B. Gottgens. 2016. Genome-scale definition of the transcriptional programme associated with compromised PU.1 activity in acute myeloid leukaemia. *Leukemia*. 30:14-23.

Slomiany, B.A., K.L. D'Arigo, M.M. Kelly, and D.T. Kurtz. 2000. C/EBPalpha inhibits cell growth via direct repression of E2F-DP-mediated transcription. *Mol Cell Biol*. 20:5986-5997.

Smale, S.T., and J.T. Kadonaga. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem*. 72:449-479.

Smith, C.C., A. Paguirigan, G.R. Jeschke, K.C. Lin, E. Massi, T. Tarver, C.S. Chin, S. Asthana, A. Olshen, K.J. Travers, S. Wang, M.J. Levis, A.E. Perl, J.P. Radich, and N.P. Shah. 2017. Heterogeneous resistance to quizartinib in acute myeloid leukemia revealed by single-cell analysis. *Blood*. 130:48-58.

Smits, E.L., S. Anguille, and Z.N. Berneman. 2013. Interferon alpha may be back on track to treat acute myeloid leukemia. *Oncoimmunology*. 2:e23619.

Soneson, C., and M.D. Robinson. 2018. Bias, robustness and scalability in single-cell differential expression analysis. *Nat Methods*. 15:255-261.

Soucek, L., J. Whitfield, C.P. Martins, A.J. Finch, D.J. Murphy, N.M. Sodir, A.N. Karnezis, L.B. Swigart, S. Nasi, and G.I. Evan. 2008. Modelling Myc inhibition as a cancer therapy. *Nature*. 455:679-683.

Spencer, D.H., M.A. Young, T.L. Lamprecht, N.M. Helton, R. Fulton, M. O'Laughlin, C. Fronick, V. Magrini, R.T. Demeter, C.A. Miller, J.M. Klco, R.K. Wilson, and T.J. Ley. 2015. Epigenomic analysis of the HOX gene loci reveals mechanisms that may control canonical expression patterns in AML and normal hematopoietic cells. *Leukemia*. 29:1279-1289.

Sproston, N.R., M. El Mohtadi, M. Slevin, W. Gilmore, and J.J. Ashworth. 2018. The Effect of C-Reactive Protein Isoforms on Nitric Oxide Production by U937 Monocytes/Macrophages. *Front Immunol*. 9:1500.

Stadhouders, R., G.J. Filion, and T. Graf. 2019. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature*. 569:345-354.

Stadhouders, R., E. Vidal, F. Serra, B. Di Stefano, F. Le Dily, J. Quilez, A. Gomez, S. Collombet, C. Berenguer, Y. Cuartero, J. Hecht, G.J. Filion, M. Beato, M.A. Marti-Renom, and T. Graf. 2018. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat Genet*. 50:238-249.

Steensma, D.P., R. Bejar, S. Jaiswal, R.C. Lindsley, M.A. Sekeres, R.P. Hasserjian, and B.L. Ebert. 2015. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 126:9-16.

Stein, S., M.G. Ott, S. Schultze-Strasser, A. Jauch, B. Burwinkel, A. Kinner, M. Schmidt, A. Kramer, J. Schwable, H. Glimm, U. Koehl, C. Preiss, C. Ball, H. Martin, G. Gohring, K. Schwarzwaelder, W.K. Hofmann, K. Karakaya, S. Tchatchou, R. Yang, P. Reinecke, K. Kuhlcke, B. Schlegelberger, A.J. Thrasher, D. Hoelzer, R. Seger, C. von Kalle, and M. Grez. 2010. Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. *Nat Med*. 16:198-204.

Stoneley, M., and A.E. Willis. 2004. Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene*. 23:3200-3207.

Su, L., Y. Tan, H. Lin, X. Liu, L. Yu, Y. Yang, S. Liu, O. Bai, Y. Yang, F. Jin, J. Sun, C. Liu, Q. Liu, S. Gao, and W. Li. 2018. Mutational spectrum of acute myeloid leukemia patients with double CEBPA mutations based on next-generation sequencing and its prognostic significance. *Oncotarget*. 9:24970-24979.

Suter, D.M., N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef. 2011. Mammalian genes are transcribed with widely different bursting kinetics. *Science*. 332:472-474.

Swiecki, M., and M. Colonna. 2015. The multifaceted biology of plasmacytoid dendritic cells. *Nat Rev Immunol*. 15:471-485.

Takahashi, K., S. Asabe, S. Wieland, U. Garaigorta, P. Gastaminza, M. Isogawa, and F.V. Chisari. 2010. Plasmacytoid dendritic cells sense hepatitis C virus-infected cells, produce interferon, and inhibit infection. *Proc Natl Acad Sci U S A*. 107:7431-7436.

Takahashi, K., and S. Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 126:663-676.

Tamura, T., P. Tailor, K. Yamaoka, H.J. Kong, H. Tsujimura, J.J. O'Shea, H. Singh, and K. Ozato. 2005. IFN regulatory factor-4 and -8 govern dendritic cell subset development and their functional diversity. *J Immunol*. 174:2573-2581.

Tang, M., J. Diao, H. Gu, I. Khatri, J. Zhao, and M.S. Cattral. 2015. Toll-like Receptor 2 Activation Promotes Tumor Dendritic Cell Dysfunction by Regulating IL-6 and IL-10 Receptor Signaling. *Cell Rep*. 13:2851-2864.

Tarazona, S., F. Garcia-Alcalde, J. Dopazo, A. Ferrer, and A. Conesa. 2011. Differential expression in RNA-seq: a matter of depth. *Genome Res*. 21:2213-2223.

Taussig, D.C., J. Vargaftig, F. Miraki-Moud, E. Griessinger, K. Sharrock, T. Luke, D. Lillington, H. Oakervee, J. Cavenagh, S.G. Agrawal, T.A. Lister, J.G. Gribben, and D. Bonnet. 2010. Leukemia-

initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction. *Blood*. 115:1976-1984.

Tawana, K., J. Wang, A. Renneville, C. Bodor, R. Hills, C. Loveday, A. Savic, F.W. Van Delft, J. Treleaven, P. Georgiades, E. Uglow, N. Asou, N. Uike, M. Debeljak, J. Jazbec, P. Ancliff, R. Gale, X. Thomas, V. Mialou, K. Dohner, L. Bullinger, B. Mueller, T. Pabst, M. Stelljes, B. Schlegelberger, E. Wozniak, S. Iqbal, J. Okosun, S. Araf, A.K. Frank, F.B. Lauridsen, B. Porse, C. Nerlov, C. Owen, I. Dokal, J. Gribben, M. Smith, C. Preudhomme, C. Chelala, J. Cavenagh, and J. Fitzgibbon. 2015. Disease evolution and outcomes in familial AML with germline CEBPA mutations. *Blood*. 126:1214-1223.

Tenen, D.G. 2003. Disruption of differentiation in human cancer: AML shows the way. *Nat Rev Cancer*. 3:89-101.

Tesfatsion, D.A. 2014. Dendritic cell vaccine against leukemia: advances and perspectives. *Immunotherapy*. 6:485-496.

Till, J.E., and C.E. Mc. 1961. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat Res*. 14:213-222.

Timchenko, N.A., M. Wilde, M. Nakanishi, J.R. Smith, and G.J. Darlington. 1996. CCAAT/enhancer-binding protein alpha (C/EBP alpha) inhibits cell proliferation through the p21 (WAF-1/CIP-1/SDI-1) protein. *Genes Dev*. 10:804-815.

Tober, J., K.E. McGrath, and J. Palis. 2008. Primitive erythropoiesis and megakaryopoiesis in the yolk sac are independent of c-myb. *Blood*. 111:2636-2639.

Toda, M., Z. Shao, K.D. Yamaguchi, T. Takagi, C.N. D'Alessandro-Gabazza, O. Taguchi, H. Salamon, L.L. Leung, E.C. Gabazza, and J. Morser. 2013. Differential gene expression in thrombomodulin (TM; CD141)(+) and TM(-) dendritic cell subsets. *PLoS One*. 8:e72392.

Togami, K., J. Kitaura, T. Uchida, D. Inoue, K. Nishimura, K.C. Kawabata, R. Nagase, S. Horikawa, K. Izawa, T. Fukuyama, F. Nakahara, T. Oki, Y. Harada, H. Harada, H. Aburatani, and T. Kitamura. 2015. A C-terminal mutant of CCAAT-enhancer-binding protein alpha (C/EBPalpha-Cm) downregulates Csf1r, a potent accelerator in the progression of acute myeloid leukemia with C/EBPalpha-Cm. *Exp Hematol*. 43:300-308 e301.

Tothova, Z., J.M. Krill-Burger, K.D. Popova, C.C. Landers, Q.L. Sievers, D. Yudovich, R. Belizaire, J.C. Aster, E.A. Morgan, A. Tsherniak, and B.L. Ebert. 2017. Multiplex CRISPR/Cas9-Based Genome Editing in Human Hematopoietic Stem Cells Models Clonal Hematopoiesis and Myeloid Neoplasia. *Cell Stem Cell*. 21:547-555 e548.

Truong, B.T., Y.J. Lee, T.A. Lodie, D.J. Park, D. Perrotti, N. Watanabe, H.P. Koeffler, H. Nakajima, D.G. Tenen, and S.C. Kogan. 2003. CCAAT/Enhancer binding proteins repress the leukemic phenotype of acute myeloid leukemia. *Blood*. 101:1141-1148.

Turner, D.L., and C.L. Cepko. 1987. A common progenitor for neurons and glia persists in rat retina late in development. *Nature*. 328:131-136.

Tzelepis, K., H. Koike-Yusa, E. De Braekeleer, Y. Li, E. Metzakopian, O.M. Dovey, A. Mupo, V. Grinkevich, M. Li, M. Mazan, M. Gozdecka, S. Ohnishi, J. Cooper, M. Patel, T. McKerrell, B. Chen, A.F. Domingues, P. Gallipoli, S. Teichmann, H. Ponstingl, U. McDermott, J. Saez-Rodriguez, B.J.P. Huntly, F. Iorio, C. Pina, G.S. Vassiliou, and K. Yusa. 2016. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep*. 17:1193-1205.

Unternaehrer, J.J., A. Chow, M. Pypaert, K. Inaba, and I. Mellman. 2007. The tetraspanin CD9 mediates lateral association of MHC class II molecules on the dendritic cell surface. *Proc Natl Acad Sci U S A*. 104:234-239.

Uttarkar, S., J. Frampton, and K.H. Klempnauer. 2017. Targeting the transcription factor Myb by small-

molecule inhibitors. *Exp Hematol*. 47:31-35.

van den Boom, V., M. Rozenveld-Geugien, F. Bonardi, D. Malanga, D. van Gosliga, A.M. Heijink, G. Viglietto, G. Morrone, F. Fusetti, E. Vellenga, and J.J. Schuringa. 2013. Nonredundant and locus-specific gene repression functions of PRC1 paralog family members in human hematopoietic stem/progenitor cells. *Blood*. 121:2452-2461.

van der Maaten, L., and G. Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research*. 9:2579-2605.

van Galen, P., V. Hovestadt, M.H. Wadsworth Ii, T.K. Hughes, G.K. Griffin, S. Battaglia, J.A. Verga, J. Stephansky, T.J. Pastika, J. Lombardi Story, G.S. Pinkus, O. Pozdnyakova, I. Galinsky, R.M. Stone, T.A. Graubert, A.K. Shalek, J.C. Aster, A.A. Lane, and B.E. Bernstein. 2019. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell*. 176:1265-1281 e1224.

Vardiman, J.W., N.L. Harris, and R.D. Brunning. 2002. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood*. 100:2292-2302.

Vardiman, J.W., J. Thiele, D.A. Arber, R.D. Brunning, M.J. Borowitz, A. Porwit, N.L. Harris, M.M. Le Beau, E. Hellstrom-Lindberg, A. Tefferi, and C.D. Bloomfield. 2009. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood*. 114:937-951.

Velten, L., S.F. Haas, S. Raffel, S. Blaszkiewicz, S. Islam, B.P. Hennig, C. Hirche, C. Lutz, E.C. Buss, D. Nowak, T. Boch, W.K. Hofmann, A.D. Ho, W. Huber, A. Trumpp, M.A. Essers, and L.M. Steinmetz. 2017. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol*. 19:271-281.

Venkatraman, A., G. Hochart, D. Bonnel, J. Stauber, S. Shimmura, L. Rajamani, K. Pervushin, and J.S. Mehta. 2019. Matrix-Assisted Laser Desorption Ionization Mass Spectrometry Imaging of Key Proteins in Corneal Samples from Lattice Dystrophy Patients with TGFBI-H626R and TGFBI-R124C Mutations. *Proteomics Clin Appl*. 13:e1800053.

Villani, A.C., R. Satija, G. Reynolds, S. Sarkizova, K. Shekhar, J. Fletcher, M. Griesbeck, A. Butler, S. Zheng, S. Lazo, L. Jardine, D. Dixon, E. Stephenson, E. Nilsson, I. Grundberg, D. McDonald, A. Filby, W. Li, P.L. De Jager, O. Rozenblatt-Rosen, A.A. Lane, M. Haniffa, A. Regev, and N. Hacohen. 2017. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*. 356.

Villar, D., P. Flicek, and D.T. Odom. 2014. Evolution of transcription factor binding in metazoans - mechanisms and functional implications. *Nat Rev Genet*. 15:221-233.

Vitali, C., C. Bassani, C. Chiodoni, E. Fellini, C. Guarnotta, S. Miotti, S. Sangaletti, F. Fuligni, L. De Cecco, P.P. Piccaluga, M.P. Colombo, and C. Tripodo. 2015. SOCS2 Controls Proliferation and Stemness of Hematopoietic Cells under Stress Conditions and Its Deregulation Marks Unfavorable Acute Leukemias. *Cancer Res*. 75:2387-2399.

Wakita, S., H. Yamaguchi, T. Ueki, K. Usuki, S. Kurosawa, Y. Kobayashi, E. Kawata, K. Tajika, S. Gomi, M. Koizumi, Y. Fujiwara, S. Yui, K. Fukunaga, T. Ryotokuji, T. Hirakawa, K. Arai, T. Kitano, F. Kosaka, H. Tamai, K. Nakayama, T. Fukuda, and K. Inokuchi. 2016. Complex molecular genetic abnormalities involving three or more genetic mutations are important prognostic factors for acute myeloid leukemia. *Leukemia*. 30:545-554.

Wan, C.K., P. Li, R. Spolski, J. Oh, A.B. Andraski, N. Du, Z.X. Yu, C.P. Dillon, D.R. Green, and W.J. Leonard. 2015. IL-21-mediated non-canonical pathway for IL-1beta production in conventional dendritic cells. *Nat Commun*. 6:7988.

Wan, Y., T.W. Kim, M. Yu, H. Zhou, M. Yamashita, Z. Kang, W. Yin, J.A. Wang, J. Thomas, G.C. Sen, G.R. Stark, and X. Li. 2011. The dual functions of IL-1 receptor-associated kinase 2 in TLR9-mediated

IFN and proinflammatory cytokine production. *J Immunol*. 186:3006-3014.

Wang, C., J.K. Sandling, N. Hagberg, O. Berggren, S. Sigurdsson, O. Karlberg, L. Ronnblom, M.L. Eloranta, and A.C. Syvanen. 2013. Genome-wide profiling of target genes for the systemic lupus erythematosus-associated transcription factors IRF5 and STAT4. *Ann Rheum Dis*. 72:96-103.

Wang, D., X.F. Huang, B. Hong, X.T. Song, L. Hu, M. Jiang, B. Zhang, H. Ning, Y. Li, C. Xu, X. Lou, B. Li, Z. Yu, J. Hu, J. Chen, F. Yang, H. Gao, G. Ding, L. Liao, L. Rollins, L. Jones, S.Y. Chen, and H. Chen. 2018. Efficacy of intracellular immune checkpoint-silenced DC vaccine. *JCI Insight*. 3.

Wang, G.G., K.R. Calvo, M.P. Pasillas, D.B. Sykes, H. Hacker, and M.P. Kamps. 2006. Quantitative production of macrophages or neutrophils ex vivo using conditional Hoxb8. *Nat Methods*. 3:287-293.

Wang, H., P. Iakova, M. Wilde, A. Welm, T. Goode, W.J. Roesler, and N.A. Timchenko. 2001. C/EBPalpha arrests cell proliferation through direct inhibition of Cdk2 and Cdk4. *Mol Cell*. 8:817-828.

Wei, G., B.J. Abraham, R. Yagi, R. Jothi, K. Cui, S. Sharma, L. Narlikar, D.L. Northrup, Q. Tang, W.E. Paul, J. Zhu, and K. Zhao. 2011. Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. *Immunity*. 35:299-311.

Weinreb, C., S. Wolock, and A.M. Klein. 2018. SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics*. 34:1246-1248.

Weissman, I.L., D.J. Anderson, and F. Gage. 2001. Stem and progenitor cells: origins, phenotypes, lineage commitments, and transdifferentiations. *Annu Rev Cell Dev Biol*. 17:387-403.

Welner, R.S., D. Bararia, G. Amabile, A. Czibere, T. Benoukraf, C. Bach, K.D. Wansa, M. Ye, H. Zhang, T. Iino, C.J. Hetherington, K. Akashi, and D.G. Tenen. 2013. C/EBPalpha is required for development of dendritic cell progenitors. *Blood*. 121:4073-4081.

Wen, X.M., J.B. Hu, J. Yang, W. Qian, D.M. Yao, Z.Q. Deng, Y.Y. Zhang, X.W. Zhu, H. Guo, J. Lin, and J. Qian. 2015. CEBPA methylation and mutation in myelodysplastic syndrome. *Med Oncol*. 32:192.

Wiemels, J.L., Z. Xiao, P.A. Buffler, A.T. Maia, X. Ma, B.M. Dicks, M.T. Smith, L. Zhang, J. Feusner, J. Wiencke, K. Pritchard-Jones, H. Kempski, and M. Greaves. 2002. In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *Blood*. 99:3801-3805.

Wilkinson, A.C., H. Nakauchi, and B. Gottgens. 2017. Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity. *Cell Syst*. 5:319-331.

Wille, S., A. Szekeres, O. Majdic, E. Prager, G. Staffler, J. Stockl, D. Kunthalert, E.E. Prieschl, T. Baumruker, H. Burtscher, G.J. Zlabinger, W. Knapp, and H. Stockinger. 2001. Characterization of CDw92 as a member of the choline transporter-like protein family regulated specifically on dendritic cells. *J Immunol*. 167:5795-5804.

Wilson, N.K., S.D. Foster, X. Wang, K. Knezevic, J. Schutte, P. Kaimakis, P.M. Chilarska, S. Kinston, W.H. Ouwehand, E. Dzierzak, J.E. Pimanda, M.F. de Bruijn, and B. Gottgens. 2010. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*. 7:532-544.

Wilson, N.K., D. Miranda-Saavedra, S. Kinston, N. Bonadies, S.D. Foster, F. Calero-Nieto, M.A. Dawson, I.J. Donaldson, S. Dumon, J. Frampton, R. Janky, X.H. Sun, S.A. Teichmann, A.J. Bannister, and B. Gottgens. 2009. The transcriptional program controlled by the stem cell leukemia gene Scl/Tal1 during early embryonic hematopoietic development. *Blood*. 113:5456-5465.

Wimmers, F., N. Subedi, N. van Buuringen, D. Heister, J. Vivie, I. Beeren-Reinieren, R. Woestenenk, H. Dolstra, A. Piruska, J.F.M. Jacobs, A. van Oudenaarden, C.G. Figdor, W.T.S. Huck, I.J.M. de Vries, and J. Tel. 2018. Single-cell analysis reveals that stochasticity and paracrine signaling control interferon-alpha production by plasmacytoid dendritic cells. *Nat Commun*. 9:3317.

Wouters, B.J., B. Lowenberg, C.A. Erpelinck-Verschueren, W.L. van Putten, P.J. Valk, and R. Delwel. 2009. Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood*. 113:3088-3091.

Wu, S., K. Li, Y. Li, T. Zhao, T. Li, Y.F. Yang, and W. Qian. 2017. Independent regulation of gene expression level and noise by histone modifications. *PLoS Comput Biol*. 13:e1005585.

Xiang, G., C.A. Keller, B. Giardine, L. An, R.C. Hardison, and Y. Zhang. 2018. S3norm: simultaneous normalization of sequencing depth and signal-to-noise ratio in epigenomic data. *bioRxiv*:506634.

Xie, H., M. Ye, R. Feng, and T. Graf. 2004. Stepwise reprogramming of B cells into macrophages. *Cell*. 117:663-676.

Yamamoto, R., Y. Morita, J. Ooehara, S. Hamanaka, M. Onodera, K.L. Rudolph, H. Ema, and H. Nakauchi. 2013. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell*. 154:1112-1126.

Yan, L., J. Ma, Y. Wang, J. Zan, Z. Wang, Y. Zhu, Y. Zhu, L. Ling, L. Cao, X. Liu, S. Li, L. Xu, Z. Qi, L. Nie, and Y. Zhang. 2018. miR-21-5p induces cell proliferation by targeting TGFBI in non-small cell lung cancer cells. *Exp Ther Med*. 16:4655-4663.

Yang, L., D. Bryder, J. Adolfsson, J. Nygren, R. Mansson, M. Sigvardsson, and S.E. Jacobsen. 2005. Identification of Lin(-)Sca1(+)kit(+)CD34(+)Flt3- short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients. *Blood*. 105:2717-2723.

Yates, J.W., H.J. Wallace, Jr., R.R. Ellison, and J.F. Holland. 1973. Cytosine arabinoside (NSC-63878) and daunorubicin (NSC-83142) therapy in acute nonlymphocytic leukemia. *Cancer Chemother Rep*. 57:485-488.

Ye, M., H. Zhang, G. Amabile, H. Yang, P.B. Staber, P. Zhang, E. Levantini, M. Alberich-Jorda, J. Zhang, A. Kawasaki, and D.G. Tenen. 2013. C/EBPa controls acquisition and maintenance of adult haematopoietic stem cell quiescence. *Nat Cell Biol*. 15:385-394.

Ye, Q., J. Jiang, G. Zhan, W. Yan, L. Huang, Y. Hu, H. Su, Q. Tong, M. Yue, H. Li, G. Yao, Y. Zhang, and H. Liu. 2016. Small molecule activation of NOTCH signalling inhibits acute myeloid leukemia. *Sci Rep*. 626510.

Yeamans, C., D. Wang, I. Paz-Priel, B.E. Torbett, D.G. Tenen, and A.D. Friedman. 2007. C/EBPalpha binds and activates the PU.1 distal enhancer to induce monocyte lineage commitment. *Blood*. 110:3136-3142.

Yin, X., H. Yu, X. Jin, J. Li, H. Guo, Q. Shi, Z. Yin, Y. Xu, X. Wang, R. Liu, S. Wang, and L. Zhang. 2017. Human Blood CD1c+ Dendritic Cells Encompass CD5high and CD5low Subsets That Differ Significantly in Phenotype, Gene Expression, and Functions. *J Immunol*. 198:1553-1564.

Yuan, H., M. Yan, G. Zhang, W. Liu, C. Deng, G. Liao, L. Xu, T. Luo, H. Yan, Z. Long, A. Shi, T. Zhao, Y. Xiao, and X. Li. 2019. CancerSEA: a cancer single-cell state atlas. *Nucleic Acids Res*. 47:D900-D908.

Zeng, X., A.M. Herndon, and J.C. Hu. 1997. Buried asparagines determine the dimerization specificities of leucine zipper mutants. *Proc Natl Acad Sci U S A*. 94:3673-3678.

Zhang, D.E., P. Zhang, N.D. Wang, C.J. Hetherington, G.J. Darlington, and D.G. Tenen. 1997. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. *Proc Natl Acad Sci U S A*. 94:569-574.

Zhang, H., M. Alberich-Jorda, G. Amabile, H. Yang, P.B. Staber, A. Di Ruscio, R.S. Welner, A. Ebralidze, J. Zhang, E. Levantini, V. Lefebvre, P.J. Valk, R. Delwel, M. Hoogenkamp, C. Nerlov, J. Cammenga, B. Saez, D.T. Scadden, C. Bonifer, M. Ye, and D.G. Tenen. 2013. Sox4 is a key

oncogenic target in C/EBPalpha mutant acute myeloid leukemia. *Cancer Cell*. 24:575-588.

Zhang, H., J.D. Gregorio, T. Iwahori, X. Zhang, O. Choi, L.L. Tolentino, T. Prestwood, Y. Carmi, and E.G. Engleman. 2017. A distinct subset of plasmacytoid dendritic cells induces activation and differentiation of B and T lymphocytes. *Proc Natl Acad Sci U S A*. 114:1988-1993.

Zhang, P., J. Iwasaki-Arai, H. Iwasaki, M.L. Fenyus, T. Dayaram, B.M. Owens, H. Shigematsu, E. Levantini, C.S. Huettner, J.A. Lekstrom-Himes, K. Akashi, and D.G. Tenen. 2004. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. *Immunity*. 21:853-863.

Zhang, W., D.E. Geiman, J.M. Shields, D.T. Dang, C.S. Mahatan, K.H. Kaestner, J.R. Biggs, A.S. Kraft, and V.W. Yang. 2000. The gut-enriched Kruppel-like factor (Kruppel-like factor 4) mediates the transactivating effect of p53 on the p21WAF1/Cip1 promoter. *J Biol Chem*. 275:18391-18398.

Zhang, X., T. Li, F. Liu, Y. Chen, J. Yao, Z. Li, Y. Huang, and J. Wang. 2019. Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. *Mol Cell*. 73:130-142 e135.

Zhang, Y., T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, and X.S. Liu. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 9:R137.

Zhong, J., P. Yang, K. Muta, R. Dong, M. Marrero, F. Gong, and C.Y. Wang. 2010. Loss of Jak2 selectively suppresses DC-mediated innate immune response and protects mice from lethal dose of LPS-induced septic shock. *PLoS One*. 5:e9593.

Zhu, B.M., K. Kang, J.H. Yu, W. Chen, H.E. Smith, D. Lee, H.W. Sun, L. Wei, and L. Hennighausen. 2012. Genome-wide analyses reveal the extent of opportunistic STAT5 binding that does not yield transcriptional activation of neighboring genes. *Nucleic Acids Res*. 40:4461-4472.

Zhu, J., X. Chen, Z. Liao, C. He, and X. Hu. 2015. TGFBI protein high expression predicts poor prognosis in colorectal cancer patients. *Int J Clin Exp Pathol*. 8:702-710.

Zhu, Q., S. Shah, R. Dries, L. Cai, and G.C. Yuan. 2018. Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. *Nat Biotechnol*.

Ziegenhain, C., B. Vieth, S. Parekh, B. Reinius, A. Guillaumet-Adkins, M. Smets, H. Leonhardt, H. Heyn, I. Hellmann, and W. Enard. 2017. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol Cell*. 65:631-643 e634.

Zuber, J., J. Shi, E. Wang, A.R. Rappaport, H. Herrmann, E.A. Sison, D. Magoon, J. Qi, K. Blatt, M. Wunderlich, M.J. Taylor, C. Johns, A. Chicas, J.C. Mulloy, S.C. Kogan, P. Brown, P. Valent, J.E. Bradner, S.W. Lowe, and C.R. Vakoc. 2011. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature*. 478:524-528.

Appendix

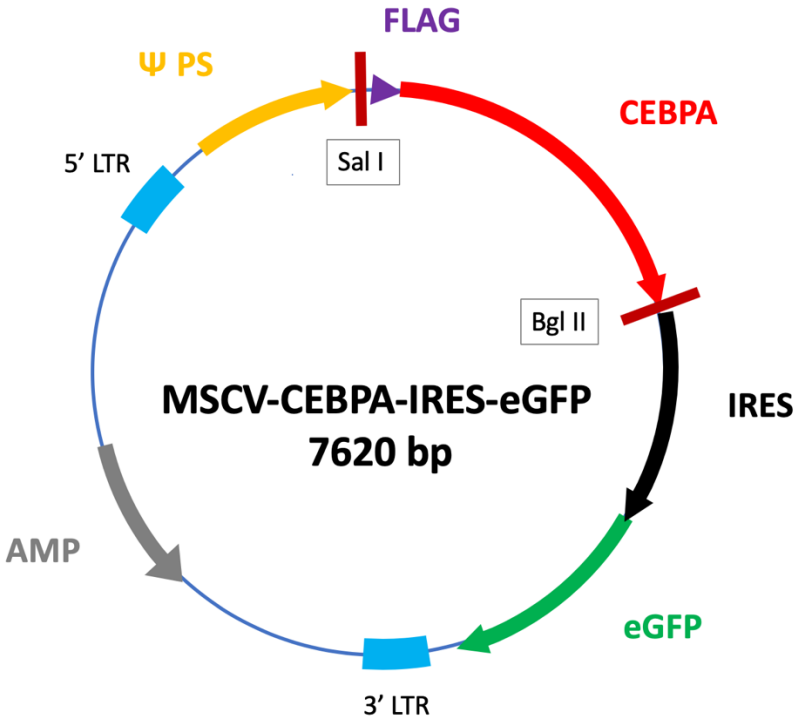


Figure 1: Vector map for MSCV-CEBPA-IRES-eGFP. Sal I (1417), FLAG (1429), CEBPA (1456), Bgl II (2533), IRES (2562), eGFP (3137).

Table 1: 636 differentially upregulated genes on Day 2 *CEBPA* WT vs Day 2 EV-transduced cells comparison

Chit1	Naip5	Acadsb	Stim1
1700047M11Rik	Far2	5730508B09Rik	Gns
Ly6c2	Ccdc125	Fam126b	Tmx3
I830127L07Rik	Zscan29	Afap1	Itm2b
Sgms2	Smurf1	Dhrs7	Cdkn2c
Tnfsf13b	Gda	Trim12a	Dynlt3
Slc40a1	Tmem216	Sun2	Ipo5
CD177	Gstm2-ps1	Dgat1	Arhgdib
4930438A08Rik	Kyat1	Gbe1	Tmem131
Abca9	Arhgef12	Trappc10	Tank
Mogat2	Flot2	Dmxl1	Hist1h1c
Tnfaip6	Nrg1	Gm5251	Cyb5r1
Tlr7	Gstm1	Arsb	Atp2b1
Cd36	S100a11	Wipi1	Gpd2
Cybb	Per2	Pmm1	Man2b2
Hsd11b1	Ugp2	Rpia	Arhgef3
Mgst1	Phlda1	Cnn2	Cspp1
Mgam	Wwp1	Bnip3l	Gng5
Chil1	Trem3	Neu1	D930015E06Rik
Fcnb	Gm13416	Pglyrp1	Map2k1
Aff2	Klf7	Chst12	Myl12b
Ngp	Prkcb	Pag1	Gm21738
Acpp	Slc38a5	Ralgapa1	Ebpl
Dgat2	Grina	Fanca	Apbb1ip

Lcn2
Abca13
Xdh
Slfn4
Fmo5
Clec4a2
Ffar2
Dram1
Gm9733
C3
Mtus1
B4galt6
Ccno
S100a8
Gpr84
Rhou
Gm29064
Tmem45a2
Lyz2
AI839979
Klra2
Elane
Draxin
Mgst2
Hcar2
Olfml2b
Itgam
Fcor
Peli2
Ltb4r1
Grb14
Ctsh
Alcam
Casp4
Mansc1
Slc44a4
Mgll
Hp
Lyz1
Igsf6
Olr1
Ncam1
Slc7a11
Ccdc109b
Wfdc21
Ptges
Gm5960
Serpinb
Napepld
Vcan
serpinb10
Rpgr
Fcgr4
Csf1
Tenm4
Gng7
Ifi206
Crisp3
Irak2
Ccl6
Csf1r
Fyb
Gm43858
Plekhg1
Aldh3b1
Ifi211
Tgm3
1700020L24Rik
Lrrk2
Scnn1a
Cd300lf
Arhgef11
Clec2i
AA474408
Plpp6
Itgb2
Lgr4
Acsl1
Gadd45a
Dock7
Nhsl2
Nrp2
Gm12854
Tmem62
Sp140
Ago3
Gm7665
Entpd1
Hlx
Syne2
Usp2
Mrgpra2b
Auts2
Pygl
Fosl2
Sort1
Gm37354
Trp53inp2
Naip2
Ascl2
Dclre1c
Mfge8
A530040E14Rik
Sorl1
Hpse
Slc31a2
Rogdi
Tnfaip2
Gm14699
Miga1
Fcgr3
Mctp1
Prdx5
Tfec
Megf9
Cyth3
Mtm1
Arl15
Gm21975
Hexb
Trib1
Rap2a
Gm37472
Cep162
Ypel5
Plekha1
Amn1
Fam134b
Glrx
Scp2-ps2
Tcn2
Soat1
Ikbke
Slco4a1
Msrb1
Gm15466
Inpp4a
Hdac4
Cpq
Svip
Ero1lb
Map1lc3a
Acta2
Atxn1
B230208H11Rik
Arhgef6
Pcyox1
Rsbn1
Mocs1
Plpp5
Fcgr2b
Gsr
Mpp1
Tmx4
Arhgap31
Stom
Ctsb
Atrn
Vps13d
Mbd5
Sri
Csf3r
Zfp622
Bnip3l-ps
Tbc1d20
Mgat4a
Kif13b
Stx6
Pnpla7
Cnbd2
4930523C07Rik
Ppp1cb
Dnajc13
Sgpp1
Irak3
PdCD4
Rab32
Gm38236
Irf2bp2
Osgin2
Osbp
Lrrfip2
CD1d2
Degs1
Mir6236
Crat
CD1d1
Samd8
Pld3
Plaur
Plgrkt
Papss2
Mcee
Dgkg
Tst
Insr
Cep85l
Sptlc2
Tmem156
Tmed3
Vps50
Lamp2
Gm5914
Zfp36l1
Gm14032
Ncf2
Tom1
AC149090.1
Cast
Stat4
Atp6v1e1
Gpcpd1
Nucb2
Cklf
Gm9025
Nfe2l2
Atp2c1
Ints12
Serpinb1a
Snx18
Ndufaf7
Gpsm3
Med10
Ftsj1
Rnf14
Mlx
Grn
Snx20
Chd7
Flot1
Prtn3
Ak3
G6pd2
Adipor1
Rock2
Gm10800
Gm26809
Tmbim6
Fuca1
Gtpbp1
Stat3
Fam46a
Serpinb1b
Gm5590
Klf13
Rpe
Map1lc3b
Smim10l1
Abtb1
Rassf5
Gm15776
Ppp1r12a
Tmsb4x
Sh2b1
Cyba
Rnaset2a
Ankle2
Apmap
Gca
Fli1
Kmt5a
Hcst
Rcsd1
Tbrg1
Zfand5
D16Ertd472e
Ankrd11
Baz2b
Ethe1
Gm33707
Arl11
Plin3
Ero1l
Golim4
Rgs2
Faap20
Bhlhe41
Arhgap26
Gm14857
Atp8b4
Chmp5
Plac8
Tnrc6b
Flna
Micu1
Dusp6
Mbnl1
Tet2
Snapin
Rbms1
Ccl3
Ifngr1

Slpi
Smkr-ps
Map4k3
Dstn
Tbc1d8
Rdh12
Syne1
Cd274
Lbp
Pabpc1l
Homer1
Atp8a1
Mmp8
F630028O10Rik
Lmbr1
Abcc5
Pld1
A430072P03Rik
Mir7678
Pirb
Anxa1
Fam160a2
Zbtb11os1
Abcd2
Ifitm6
Fgl2
Zdhhc2
Hsd17b1
Abhd5
CD14
Cldn15
Nfil3
Ccm2l
Lgals3
Cmpk2
Vsir
Mocos
Celsr3
Slc26a6
Gas7
Sqrdl
Csf2rb2
Ceacam1
Gm16139
Znfx1
S1pr1
Camsap2
Pex26
Rflnb
Dach1
Aoah
Maml1
B430010I23Rik
Gfi1
Cyp4v3
Pi16
Sash1
Serpinb6a
Tlr4
Slc8b1
Afmid
Serpinb10

Mospd1
Pnkp
Per3
Bmx
Gm4854
Scp2
Ppm1h
Prkd3
Slc22a4
Gm8770
Gm2147
Fnip1
Anxa3
Ptgr1
Lmtk2
1110008P14Rik
Dubr
Btg1
Lta4h
Gm5552
Fgd4
Aplp2
Gyg
Agps
Ap1s2
Rgcc
Tmem63a
Lcmt1
Il16
Sat1
Mirt1
Tmem38b
Slc17a5
Aldh3a2
Mpp7
Vsig10l
Mafg
1700123O20Rik
Tmem159
Tle3
Alas1
Mpo
Dhrs1
Gm14005
Pisd-ps1
Stambpl1
Sgpl1
Acox3
Pon3
Csf2rb
Fam21
Gapt
Dync2h1
Daam1
1810037I17Rik
Pecr
Sp100
Acad11
Ginm1
Lyst
Gm3150
Tma16

Cd24a
Micu3
Bend4
Lpgat1
Cbr4
Plin2
Cyb5r4
Tcp11l2
Slc25a37
Csgalnact2
Ankrd28
Rnase4
Gm5884
Spi1
Agtrap
Elmsan1
Usp33
Ric1
Rin3
Cdadc1
Dram2
Atp6v1a
Rnaset2b
Npepps
Smim12
Gm11868
9130401M01Rik
Cd53
Gm10801
Gm1840
Phkb
Gpi1
Fam132a
Rab27a
BC005561
Tiam1
Acad8
Pam
Gm13342
G6pdx
Pxylp1
Dctn1
Sirpa
Gpr160
Sidt2
Trip4
Gys1
Bet1l
Mid1ip1
Csf2ra
Riok3
Hipk1
Aff1
Tnfaip8
Maml3
Zfp992
Edem1
Pink1
Rnf111
Vamp4
Selenop
Esf1

Bin3
Nars
Sla
Vim
Atp6v0c
Hdac5
Shfm1
Plod3
Pde6d
Rapgef6
Zyx
Zmpste24
Smim4
Adssl1
Srgn
Gm13082
Luc7l2
Cpne3
Ptprc
Selenos
Slc50a1
Ctsd
Snhg5
Epb41l4aos
Tmem165
Wls
Tor1a
Arhgap15
Alox5ap
Atp6ap2
Ndfip1
Pgm2
Ccdc50
Ncoa4
Serp1
Mbnl2
Ddx17
Tpp1
Rab6a
Gm6768
Asnsd1
Taldo1
Akap9
Cd33
Fam49b
Uggt1
Idh1
Ostf1
Rsrp1
Atp6v0b
Lin7c
Psen1
Stt3b
Atp6v0d1
Gm15487
Gm12428
Oxr1
Gmip
Wbp5
Bckdha
Gm7707
Serf2

Table 2: Enrichr GO biological processes from 636 differentially upregulated genes on Day 2 *CEBPA* WT vs Day 2 EV-transduced cells comparison

Index	Name	Adjusted p-value	Odds Ratio	Combined score
1	neutrophil degranulation (GO:0043312)	1.28E-25	4.78	314.97
2	neutrophil activation involved in immune response (GO:0002283)	1.10E-25	4.74	309.8
3	neutrophil mediated immunity (GO:0002446)	1.25E-25	4.7	304.75
4	granulocyte migration (GO:0097530)	0.02793	17.97	185.51
5	toll-like receptor 4 signaling pathway (GO:0034142)	0.04704	10.48	99.74
6	positive regulation of reactive oxygen species metabolic process (GO:2000379)	0.003628	7.08	88.78
7	pattern recognition receptor signaling pathway (GO:0002221)	0.00337	6.42	82.43
8	ribose phosphate metabolic process (GO:0019693)	0.1228	10.48	81.78
9	pentose-phosphate shunt (GO:0006098)	0.1159	10.48	81.78
10	cellular glucan metabolic process (GO:0006073)	0.1098	10.48	81

Table 3: 429 differentially downregulated genes on Day 2 *CEBPA* WT vs Day 2 EV-transduced cells comparison

Tmem268	Gfi1b	Prcc	Syce2
Ighm	Ctla2a	Ass1	Plek
Ddx4	Ppp1r13b	Gm29397	Apobec3
Gm15261	Jade2	Dnmt3b	Nrm
Il7r	Ccdc9	2610318N02Rik	Alyref2
Cdk18	Plekho1	Hist1h2ai	Cfp
Mmp13	2310075C17Rik	Cep78	Msh6
Ms4a6b	Dab2	Csnk1g2	Gm13005
Rag1	Fam65a	Slc2a6	Rpf2
Itpr3	F2rl3	Gch1	Dhfr
Pear1	Fkbp7	Socs2	Acap1
Mzb1	Chl1	Dnajc6	Cfap20
Crocc	Rps6ka3	Cnn3	Gm4849
C030034L19Rik	Cd34	Grwd1	Lgals1
Spns2	Igkv12-89	Tmem173	Atp1b3
Gata2	4930427A07Rik	Fam212a	Bin2
Mc5r	Gata3	Fen1	Smarca5-ps
Drc7	Coro2a	Akap1	Plod2
CD163	Spp1	Gm28875	H2afy
Dntt	Cerkl	Crip1	H2-DMa
Slc18a1	Cav2	Meis1	Cenph
Il21r	Vamp5	Elk3	Sord
Prr5	Usp12	Gm9847	Gm21887
Nlrp10	Psd3	Gm15654	Rfc3
D630045J12Rik	Relt	Ttll9	Gm37201
Cyp4f18	Hdac9	Mtap	Sae1
Rtn4r	Flt3	Tmpo	Gm10193
Rtp4	Car3	Chuk	Csk

Carns1
Tmem176a
Zfpm1
Simc1
Dusp4
Hrh2
Egr2
Car8
Ralgds
Fam129b
Tesc
Ttyh2
Gm38345
Pi15
Gm11998
Itgb7
Gpam
Gm16897
Col23a1
Pank1
Slc45a3
Nr1d2
Gfra2
Gfra1
Dgkq
Gm43696
Ctxn1
Afap1l1
Lsmem1
Pars2
Tmem132a
Ighv1-23
Spa17
CD180
Tspan6
Thsd1
Cd244
Il31ra
Capn3
S100a4
Mex3a
Rbp1
Gimap5
Nxpe5
Bex6
Otos
Gm21807
Arhgap6
2610307P16Rik
Etv5
Tada2b
Pbx3
AW112010
Nhsl1
Itga9
Cela1
Il6
Atp2b4
Mrgpre
Dgkh
Zfp566
Angptl2
Rasgef1b
8430408G22Rik
Zfp729a
Plscr4
Mast4
Tmem119
Mtmr10
Cd72
Tnfrsf12a
Tram2
Abcb1b
Crls1
Adgrg1
Nfya
Myl4
Gm15581
Igkv4-90
Etv4
S100a6
Muc13
Smtn
Emid1
Slc36a3os
Carm1
Ptgds
Spaca9
Praf2
Nkg7
2900026A02Rik
Golm1
Nrgn
Kcnn4
Sh2d5
Serpinf1
Spns3
Tgfb1
Irx2
Pkib
Nup35
Gm3470
Plcb4
Fbf1
Stim2
Hist1h2ag
Rbpms2
Mcm10
Adgrl4
Dscc1
Bahcc1
4930519L02Rik
Supt16
Gsk3a
Acss1
B3gnt7
H2-Ob
Gon4l
Tmem176b
Slc35d3
Flnb
Hps5
Stk10
Cnot6
Ctla2b
Ccnd1
Parp9
Lat2
D5Ertd605e
Ifi203
Dhx33
Ankrd37
Gm17494
3110009E18Rik
Tmem238
Gm5424
Peg13
Gm12470
Gm12611
Adrb2
Car2
Gm17936
Cbfa2t3
Cenpv
Lrp5
Hist1h2bj
Fam60a
Trp53i11
Trim37
Rcc1
Kti12
2700099C18Rik
2610524H06Rik
Gm6169
Zfp422
Txndc5
Bmp2k
Lgals9
Angpt1
Ssbp2
Helb
Gm15658
Pus10
Gm24452
Nasp
Il17ra
Ptprcap
Pkig
Tsc22d1
Trp53
Tarbp2
Lmnb1
Scaf8
Hmgb3
Mcpt8
Stmn1
Nmral1
Mlkl
Isoc1
Itga6
Rbl1
Cdk5rap2
Nfatc1
Gm11223
Blm
Rpa1
Ppih
Slc11a1
Rbbp7
Gm6382
Gm14769
Trp53-ps
Gm12184
Slc16a10
Gm11246
Hells
Smarcc1
Smc2
Rfc4
Steap3
Gm1848
Lsm3
Slc39a6
Topbp1
Gcsh
Clic4
Arhgap18
Mcm3
Endod1
Psmb10
Gm6104
Gm2a
Slc38a1
Gm15428
Mcm6
Gm5593
Gm5385
Zfp36l2
Fbl
Eef1akmt1
Gm5529
Nrp1
Aimp2
Parvg
Abce1
Ifitm1
Nme1
Gm42965
Gm7733
Aarsd1
Ptpn18
Gm10087
Psme3
Iars
Pgrmc1
Gm6652
Isyna1
Hist1h4i
Eif3b
Polr3h
Wdr18
Rbm38
Fkbp3
Ftsj3
Gm20156
Gm5837
Gm6640
Plp2
Gins2
Emp3
Gm5530
Kansl2-ps
Pold2
Alyref
Kansl2
Gm10167
Ranbp1
Gm43578
Gart
Shmt2
2810004N23Rik
Gm1969
Gm5124
Gm5256
Cdca7
Gm4617
Nop58
Cse1l
Med29
Cops4
Gm9800
Slc35b4
Tuba1b
Gm3687
Ly6e
Rrm1
2810417H13Rik
Nipsnap3b
Lmo2
Gm6625
Mcm4
Hmgb1-ps7
CD48
Mcm2
Higd1a
Wdr5
Mrps23
Tipin
Gar1
Gm11914
Hmgn5
Anp32-ps

Thy1	Gm6960	Dapp1	Gm7901
Lpar5	Psat1	Ykt6	Ptma
Rcor2	Tmem229b	Nup205	Prmt1
Kcna3	Rgs1	Gm13521	Rpl21
Plxnb2	Gm15773	Gins1	
Ccdc51	Knl1	Usp22	
Tespa1	Hist1h3c	Tcf3	

Table 4: Enrichr cell ontology (mouse gene atlas) of 429 differentially downregulated genes from Day 2 *CEBPA* WT vs Day 2 EV-transduced cells comparison

Index	Name	Adjusted p-value	Odds Ratio	Combined score
1	stem_cells__HSC	0.000264	4.01	51.35
2	common_myeloid_progenitor	0.03158	4.08	26.15
3	embryonic_stem_line_V26_2_p16	0.005219	2.05	18.7
4	embryonic_stem_line_Bruce4_p13	0.0104	1.86	14.96
5	bone_marrow	0.03843	2.14	13.81
6	follicular_B-cells	0.09355	1.78	9.14
7	macrophage_bone_marrow_0hr	0.4888	2.07	6.63
8	mega_erythrocyte_progenitor	0.2659	1.66	6.54
9	mast_cells_IgE+antigen_1hr	0.6459	1.76	4.93
10	nih_3T3	0.7533	1.67	3.93

Table 5: 15 differentially upregulated genes and 5 differentially downregulated on Day 2 *CEBPA* N321D vs Day 2 EV-transduced cells comparison

Upregulated			Downregulated
Slc26a6	Mmp12	Gm14328	Spns2
Mmp8	Lmo1	Gm13998	Gm16104
Ifi214	Fabp5l2	Fabp5	S100a6
AA474408	Mir6236	Lsm3	Zfp36l2
Slpi	Gm6166	Anxa3	Adgrl4

Table 6: Genes upregulated during culture in Flt3L differentiation media

151 genes differentially upregulated exclusively in *CEBPA* N321D-transduced cells

Gm2622	Igfbp4	Sfxn3	Tpm2
Luc7l2	Six1	1700025G04Rik	Mpc2
Dleu2	Dusp7	Gm17887	Serpinf1
Gm1848	Trpc4ap	Ocel1	Osbpl3
Bnip3l	Atp7a	Atxn1l	Rnf150
Scin	Tns1	Sh3bp1	D930015E06Rik
Trappc5	Klrb1c	Fli1	Gfod1
Iffo1	Snx29	Apc	Gm8615
Tmed8	Zfp947	Gm12045	Egfl7
Rprd2	Trim3	Tmem261	Dcaf10
Ncoa4	Tmem119	Ivns1abp	Serinc3
Med13l	Gm8666	Atp13a2	Shisa2
Pafah1b3	Borcs6	Zfp933	Rab5a
Nup210	Gm8752	Tmem8c	Fam65b
Lnpep	Ctns	Mylpf	Slc44a1
Tubb6	Avl9	Gm7694	Mef2c
Tmem170b	Mef2d	Myog	Klra1
Cbfa2t3	H2-M3	Tmem229b-ps	Tm6sf1
Vamp2	Fam234a	Mgat1	Tifab
Cirbp	Fabp4	Clec2i	Nceh1
Hoxb8	Ube2a	Fam214a	Gm8181
Pigv	Ssh2	Znrf1	Satb1
Map3k5	Hmgn1	Wasf2	Notch1
Adgre5	Hotairm1	Tnfrsf23	Lhfp
Atp1b3	Cox6a2	Fam134b	1110059G10Rik
Zeb2	CD163	Bloc1s2-ps	Smim20
Bcl2	Fam76b	Zfp296	Pecam1
Pink1	Fyn	Mroh1	Zbtb18
Gatsl2	C330007P06Rik	Fbxw4	Gm26798
Parp8	Cmah	D930048N14Rik	Aagab
Mtm1	Gab3	Sema6d	2310001H17Rik
Sh2b2	Rnf130	Klhl30	Gm7862
Rsrp1	Was	Dgkd	Gm11663
Hes6	Fads2	Abcb4	Bmf
D8Ertd82e	Slc27a1	Dap	Dr1
Runx2	Fndc3a	Rab33b	Cd93
Adcy7	Gm6088	Sgcg	Mtpn
Kcnk12	H2-T-ps	Dhrs3	

956 genes differentially upregulated exclusively in EV-transduced cells

Adap2	Sh3tc1	Trp53i13	Aoah
Tacc1	Myd88	Gm37347	Zyx
Med10	Wdr45b	Slc7a8	Casp4
Dram2	Ehbp1l1	Nuak2	CD80
Crybg3	Parp10	Ftl1-ps1	Gm13077
Plekhm1	Sertad3	Tbc1d8	Pea15a
Mthfs	Smchd1	H2-DMb2	Aldh3a2
Tapbpl	Lpcat4	Apoc2	Gm7823
Myo1g	Arid5a	Phf11d	Stxbp3
Adap2os	Dennd1a	Tet3	Ncoa7
Zfyve26	Tmem192	Ciart	Acap2
Spty2d1	Ifit3	Ptpra	Hexa
Hsd17b4	Csf2ra	Serp1	Gpr160
Samd9l	Gm5431	Slamf9	Slc29a3
H2-Ab1	Tmem106a	Gna15	Mknk1
Ifi30	Zdhhc7	Hmgcl	Egr1
Fuca2	Gpr155	Gm8185	Nadk
Max	Gm13597	S1pr2	Mov10
Hcar2	Erlin1	Tnfrsf14	Snx21
Rnf213	Sgk3	Cd2ap	Cln6
Cd274	Rcbtb2	S100a11	Rnf34
Cmtm3	Scarb2	Lipa	Tm2d2
Tlr9	Selplg	H2-K2	Pilrb1
Ggh	Arhgef6	Dclre1c	Lipe
Gm18588	Slc15a3	S100a6	Irak2
Trim26	Rsad2	Psme2	Arl5c
Rnf115	Mafk	Cd244	Cdk14
Lrp1	Sh3bp2	Snap23	Lmln

Prkch
Maf
Gm13392
Gm15989
Gm43245
Nrp1
Coro1a
Ifit1bl1
Gm15487
Calm1
Rnh1
Cd22
Pepd
CD48
B430306N03Rik
Spg21
Ier2
Metrnl
Tlr13
Gm13453
Snx2
Atp6v1e1
Add3
Gpr108
Gm7114
Klf10
Etv3
Carmil1
Sorl1
Clec9a
Il4ra
Ptpn22
Itpr1
Cstb
Atp2b1
Gng10
Lpp
Brox
Degs1
Smox
Rgs1
Amn1
Spaca9
1700097N02Rik
Rtn4
Capza2
9430034N14Rik
Rarg
Gbp4
Nmi
Il3ra
Erp29
Psme2b
Clcn5
Rnpep
Scn4b
Malat1
Psme1
C3
Lpar6
Ly6c1
Tfe3
Macf1
Itpripl2
Btg2
Mcl1
Lyst
Gbp9
Gm7367
Gm7776
Dab2
Fam49b
Gm13535
Mfsd5
Anxa4
Prkab2
Gm6377
Fkbp1b
Lamp2
F630028O10Rik
HaCD4
Ccnd1
L1cam
Fem1c
Gm42786
Actg1
Psen1
Gvin1
Oser1
Gm12715
P2ry14
Slc25a38
Dtx2
Slfn10-ps
Gadd45g
Irf1
Clec4n
2610206C17Rik
Slc35c1
Entpd1
Mfsd12
Aff1
Prkd3
Sft2d1
Txndc15
Cd300lg
Elmo2
Jak2
Gnaz
Ly9
Galnt10
Tnf
Stard3
Hexb
Asgr2
Slamf7
Sirpa
Phf11b
Lrch1
Gm6615
Fbxw17
Taf13
Rps6ka4
Exoc3
Ppp3ca
Gm26637
Nfatc2
Coro1b
Atp6ap1
Gpr132
Ankrd44
Rbpj
Sbno2
Gm4918
Oasl1
Ostm1
Nipa2
Smap1
Anxa2
Sema4c
Arhgap4
Fcgrt
Emp3
Ascc3
Chmp4b
Nkiras2
Zfp36l1
G6pd2
Gm13212
Vcpip1
Bcl10
Gm6977
Arl8a
Ensa
Pik3ap1
Polb
Cxcr4
Renbp
Cldnd1
Vps37b
Bco2
Zfp992
Anxa1
Gm15361
Gm5510
Gpr18
Dram1
Socs1
Nt5c3
Nfkb1
Il18
Akna
Acot9
Azi2
Ccl4
Clta
Rab3d
Lmo1
Rab32
Slc26a2
A230046K03Rik
Zfp382
Tpm4
Lrrc25
Nek7
St18
Ifi203-ps
Slc35f6
Gm4760
Haus8
Tnfaip2
Tmem176b
Lrrfip2
Ncoa1
Nub1
Pqlc3
Ppp2r5a
Cnr2
Vwa5a
Cyp4f16
Neurl3
Rras2
Tmem154
Dtx3l
Cic
Itpripl1
Sirt7
Acss1
Actg-ps1
Sgpl1
Fbrsl1
Plekhg3
Dock5
Mtf1
Necap2
Mob3a
Rn7sk
Zfand2a
Acbd5
Tnfsf12
Gm1966
Ifi214
Ppp6r1
Oas1a
Dbnl
Dck
Rasgef1b
Tmem9b
Birc3
Rraga
H2-Eb1
Morc3
Lilra6
Picalm
Gm23374
Eif2ak1
Gm15534
Zfp36
Kynu
Gm21738
Cyba
Rel
Myo1c
Ccdc109b
Junb
Lpl
Gm26740
Flna
Fnbp1
Arhgap30
Vrk1
Spop
Cd53
Stxbp3-ps
Chmp1a
Foxn2
Susd3
Trim5
Lgals8
Sp110
H2-Aa
Plxnc1
Ifit2
Sirpb1b
Fry
Cmtr1
Sri
Samd8
Hmox1
Rbms1
Gm2986
Pttg1
Ddit3
Dyrk1a
Adgre4
Mgat4a
CAAA01147332.1
Nfkbib
Igtp
Ccdc86
Kif1b
Gm5763
Acvrl1
Fam69a
Dync1h1
Phf11a
Chn2
4930455G09Rik
Hcfc1r1
Ifi47
Nampt
Gm43330

Tor1a
Hist3h2a
4833421G17Rik
Lamtor1
Lacc1
Irgm2
Gm8641
Rchy1
Tmem176a
Mndal
Szrd1
Zfp277
Aif1
Ptafr
Rgl1
Ap2a2
1600010M07Rik
N4bp2l1
Gm5828
Cap1
Clcn7
Ep300
Gm5578
Plin2
Hk2
Gm21975
Gm8577
Pfkfb4
Adora3
Mysm1
Myo9b
Bin3
Naip6
Slc34a2
Swap70
Cxcl2
Akt3
Frmd4a
Cyb561a3
1110008P14Rik
Rabep1
Vim
Hpgds
Plaur
Rnf166
Gm15899
Pisd
Ncf2
Bet1l
Plxdc1
Stk38
Il2rg
Racgap1
Plekhb2
Dusp10
Diaph2
Slc31a1
Ndel1
Herc6
Ifi213
Ppm1h
Slc8b1
Vapb
Fcgr4
Tnip3
Npc2
Sec24b
Ero1lb
Ms4a4a
Ccr2
Arrb2
Fos
Rap2c
Tmem150b
Gm8902
Zfp467
Arpc5
Trim30c
Apobec1
Spib
Psmb10
Crlf2
Gbp2
Atp6v0d1
Maml1
Fam129a
0610012G03Rik
Syk
Armc7
Hgsnat
Gm15922
Tmem86a
Gm6175
Mycbp2
Gsdmd
H2-DMb1
Ccdc88a
M6pr
Pdlim5
Gm5548
Zcchc6
Dennd1c
Pilrb2
Aim2
Raph1
Wwp1
Zcchc2
Tlr7
Azin1
St8sia1
Plcl2
Fcer1g
Gm20432
Dusp3
Rell1
Zc3h12a
Tnfsf8
Mpc1
Klk1b11
Parp12
Ccdc12
Oas1g
Gm5046
Atg7
Gm10801
2810405F17Rik
Pcsk7
Dnajb14
Acta2
P2ry12
Cib1
Ctsc
Rgs14
Tcf19
Fnip2
Epb41l2
Atp6v0b
Keap1
AW112010
Bcl3
Rassf1
Card11
Actr3
Grasp
Msr1
Usp25
Cst12
1600014C10Rik
Trim36
Ifi209
P2rx4
Lyz1
4930430E12Rik
Ncoa3
Cyld
Plpp6
Plod1
Tlr1
Vamp3
Socs3
Arhgap15
Arhgap25
Tmsb4x
Gm16205
Gm9392
3110043O21Rik
Mob1a
Lrrc75a
Mapk14
Klf6
Gabarap
Dennd1b
Cass4
Paqr4
Ralb
Dock4
Slc43a2
Plekho2
Cln3
Mvb12a
Ttyh3
Gna13
Gm12250
Gm15503
Tspyl1
Pml
Ccl2
Ifit3b
Gm12791
Klf4
Cmip
Nedd9
Usf1
Bin1
Prdx5
Fcgr1
Rap2b
Ifit1
Ifngr1
CD40
Srgap2
Dcaf11
Rgs2
Tle3
Ppp1r18
Gpr171
Aldh3b1
Arhgdia
Pstpip1
Efhd2
Itm2b
Gm8369
E2f7
Herpud1
Slfn9
Sub1
Tbc1d13
Naaa
Wipf1
Ptp4a2
Tnfrsf1a
Cxcl16
Ifngr2
Mdm2
Atp6v1b2
Ssh1
Sap30
Cttnbp2nl
Scand1
Relb
Eif2ak2
Gm8464
Ctsd
Vhl
Gm12428
Rnf114
Gm7676
Txndc16
Pias1
Tmod3
St3gal2
Gm42641
Wdr44
Phactr2
Fgl2
Tlr6
Cast
Leprot
Bcl2l11
Mdfic
H2-T24
Tmem134
Far1
Nfkb2
Tbc1d9
Foxred2
Cd9
Dpysl2
Smpdl3b
Rnd3
Trim12c
Slc15a4
Trim14
Ptpn1
CD86
Mir6236
Cdc42
M6pr-ps
Iqgap2
Cyp4f18
Slc37a2
Ptms
Rab8a
Pnkp
Lpin1
Gpr137b-ps
Plekhm3
Slc25a22
Dnase1l1
Ifi204
Arpc2
Slfn1
AC133103.1
Cpne3
Cbl
Gnai2
Ier5
Sp2
Gm6263
Pid1
Gm11405
Usf2

Twf2	Ids	9430076C15Rik	Kif23
Dhrs11	Apbb1ip	Irgm1	Gm9025
Rgs3	Ccrl2	Rnf167	Net1
Clec4a4	Gm14856	Stard9	Lactb
Dpy19l1	Clec4b1	Cd9-ps	Rin3
Ftl1	Lcp1	Creb3	Alox5
Synj1	Gngt2	Gm4750	Il12rb2
Ms4a7	B4galt5	Tnfaip8l2	2310035C23Rik
Ifi206	Gm3145	Rap1b	Vcl
Taldo1	Gm12543	Elmsan1	Mif4gd
Tmem184b	Ptk2b	Nagpa	Dpp4
Map3k3	Gm12226	Atp8a1	Osbpl9
Gm11189	Cxcl10	Thap6	Rnf31
Skap2	Rab1b	Gm10800	Malt1
Atf3	Atp6v0c	Card19	Crip1
Zbtb7b	Nxpe4	Fam26f	March5
Vps51	Pdxk	Sirt2	Skil
Zfhx3	Tmem219	Helz2	Zfp710
AC132444.1	Map2k1	Gm9397	Rala
Atp1b1	Ubc	P2ry10	Pik3cd
Acot10	Emilin2	Gm15448	Gm16026
Rab3il1	Sdhaf2	Tgfbi	Lrrk2
Sh3glb1	BC005537	Isg20	Gm4962
Gm12430	Kctd12	Slc35c2	Slc25a45
Camk1	Tax1bp3	Msrb1	Lmna
Abhd12	Snx10	Tmem189	Pim1
C3ar1	Esyt1	Mical1	Nfkbid
Commd8	Ccdc50	H2-DMa	Adgre1
Actb	Pnpla7	Ints4	Lrrk1
Taf6l	Syngr2	C130026I21Rik	Padi2
Gm8399	Lrch3	Sdccag3	Fam46c
Stat6	Il21r	Ttc7	Laptm4a
H2-T22	Gm17087	Akr1a1	Capn2
Parp3	Trerf1	Trim30b	E2f2
Actr2	Trem2	Adar	Ifih1
Adam15	Cdk13	Hpcal1	Il15
5430427O19Rik	Rnf13	Grina	Klra17
Sertad1	0610040J01Rik	Stard3nl	Gm37589
Ifi208	Ifnar1	Gm20559	Gsap
Gm7665	Acaa1b	Gm6305	Nfkbiz
Zbp1	Ptprj	Aim1	Cdc42se2
Stam2	Dna2	Sidt2	Rps6ka5
C1qc	Gltp	Arf5	Ywhah
Peli1	Npc1	Iqgap1	Nckipsd
Arhgap31	Cnn2	Tlr4	Ankfy1
Gak	Fcgr3	B3gnt8	Clec4a1
Rac1	Gm10717	Gm14000	Daxx
Ncor2	Igsf6	Hck	Gm12164
Nostrin	G6pdx	Grb2	Gm7357
Grap	Sema4a	Rbfa	Sgpp1
Ostf1	Bhlhe40	Mvp	Scimp
Wdr1	Gm12166	Ddi2	Cotl1
Dusp1	Gm15753	Pqlc1	Nfe2l2
Gm8822	Gm5590	Camta2	Fcor
Gpr146	Sqstm1	Slfn8	Upb1
Traf5	Antxr2	Fuca1	Rp2
Smdt1	Gmip	Pelo	Cav2
Icosl	Uvrag	Id3	Fgd3
Gm6695	C1qb	Ifi35	Fnip1
Sesn2	Tmem173	Hdac4	Vps37a
Xbp1	Ptger4	Gm12003	Siglec1
Gm7809	Cndp2	Rhog	Mpp1
Ccdc50-ps	Gnpda1	Fth1	Vasp
Fgr	Ltb4r1	Pot1b	Abcg1
Pag1	Map7d1	AB124611	Mgl2

443 genes differentially upregulated in both *CEBPA* N321D and EV-transduced cells

Tnfrsf13b	Selenop	Ppm1m	Csf1r
5730416F02Rik	Pmaip1	Ythdf3	Cds1
Plxnb2	Trafd1	Tubb2a	Stk17b
Slc2a6	Ifi207	Ifi205	Txnip
Siglech	Trim25	Slfn2	March1

Blvrb
Lpxn
Itgb7
Mbnl1
Gm5398
Bex6
Stat2
Tyrobp
Irf5
Uba7
Inpp4a
Mycl
Cdkn1a
Ddx58
Axl
Slfn5
Limd2
Filip1l
Ddx24
Atp6v1d
Tapbp
Camk1d
Napsa
Il17ra
Sdc4
Tlr2
Tmsb10
Arpc5l
Acaa1a
Ly6e
Plac8
Nfkbia
Ubl3
Cdc42se1
P2ry13
Pirb
Ctsh
Trim12a
Tet2
Il6st
Ms4a4c
Tor3a
Mkrn1-ps1
Cx3cr1
Epn1
Cd52
Slc37a3
Gm7936
Gng2-ps1
Fads3
Dpep2
Scpep1
Klrd1
Ptpn6
Ighj4
Fes
Tmem109
Litaf
Map3k8
Trim30a
Ckb
Rpph1
Tmem229b
Rnase6
Gm14032
Wdr26
Pik3r5
Tpst1
Man2b1
Fam105a
Cytip
Tecpr1
Csf2rb
Cybb
Myadm
Rasa4
Il1r2
Gm4875
Snx20
Gm14699
Ctsa
Cd28
Tspo
Alcam
Xiap
Arl6ip5
St8sia4
Laptm5
Myo1f
Sgcb
Ms4a6d
Gm5552
Prkcd
Gosr2
Sh3bgrl3
Tep1
Gm14548
Ptpro
Cnp
Ctse
Notch2
Sat1
Selenow
Itgax
Arhgap9
Smim5
Btg1
Cyth4
Unc93b1
Whsc1l1
Evi2a
Hmha1
Mir22hg
Mpc1-ps
BE692007
Oas3
Itgb2
Klhl6
Ifi203
Samhd1
Hlx
Usp12
Ctnna1
H1f0
2010001K21Rik
Gsn
Cfh
I830127L07Rik
Pisd-ps1
Xaf1
Abi1
Plec
Themis2
Ly6c2
Slc44a2
Itga1
Reep5
Pak1
Smim3
Bcl6
Hpse
9930111J21Rik1
Lair1
Slc9a7
Prr5
Fgd2
Gm15931
Bst2
P2ry6
Parp9
Hps3
Ccl6
Cd74
Alox5ap
Sdc3
Gng2
Tgtp2
Abhd15
Slc16a7
Arl4c
Runx1
Ccdc180
Gm15417
Cebpd
Mef2a
Acer3
Ly86
Anxa6
Capg
Havcr2
Pip4k2a
Fmnl1
Prkx
Stat1
Lilrb4a
Lbh
Plekha1
Cmtm6
Epb41l4b
Bach2
Clec4a2
Rassf5
Map3k1
Lst1
Gas7
Gpr141
Inpp5k
Clec4a3
Tec
Tmem55b
Tmem243
Neat1
Parp14
Tnni2
D1Ertd622e
Klf3
Cd300lb
Gm12345
Sh2b3
Oasl2
Gm3788
Tor1aip1
Cln8
Znfx1
Sik1
Mkrn1
Nfatc1
Rogdi
Tdrd7
Zfp385a
Ifi211
A530040E14Rik
CD14
Zfp263
Tiparp
Map4
H2-T23
Pilra
Lilr4b
Spi1
Cd300c
Gm13669
Tmed3
Kdm7a
Arhgef12
Irf9
Rgs10
Hcls1
Ighm
Usp18
Clec10a
Tpd52
Prex1
Trpv2
Ehd4
Itm2c
Cnbd2
Ube2l6
Gpr34
Msn
N4bp3
Lifr
Zc3hav1
Nptn
Lgals3bp
Ctnnd2
Fam174a
Gm14608
Cd300a
Fam129b
Arhgap17
Il6ra
Pitpna
Nfam1
Grn
Ap1s2
Slc14a1
Pld4
Tifa
Tmem50b
Rassf4
Ogfr
Kmo
Stx7
Gpsm3
Ly6a
Fxyd5
Nlrp3
Nrros
Anxa5
Lyn
Tulp4
Sp100
Ccr1
Gpcpd1
9930111J21Rik2
Milr1
Cfp
Tnfsf9
Zufsp
Lcp2
Irf7
Naga
Cysltr1
Ifnar2
Trim30d
Klrb1f
Arhgap5
Ypel3
Il13ra1
Cd300lf
Ms4a4b

Tubgcp5	Naip5	Sft2d2	Cd68
Cadm1	Csf2rb2	Gm17791	5031439G07Rik
Gm16372	Psap	Rubcnl	Nr4a1
Ypel5	Necap1	Gm8995	Oas2
St3gal5	Dhx58	Gm12892	Pycard
Gm15987	Prr13	Kctd14	Rbm47
Smim14	Mtmr14	Isg15	S100a4
Fyb	Zbtb7a	Trim34a	Il10ra
Vsir	Lgals3	Acox3	Ahnak
Lgmn	Ms4a6b	Gm12854	Bloc1s1
Wfdc17	E330020D12Rik	Ifitm3	Slc46a3
Nos1ap	AI839979	Cst3	Tpd52-ps
Rasgrp4	Lasp1	Gpr183	Sppl2a
Zfp36l2	Tmcc3	Ebi3	Ak8
Shisa5	Bmyc	Plp2	Gcnt2
CD82	Gns	Plekhg5	Tcirg1
Ccl3	Ifi27l2a	Tmem51	Ighj2
Rab7b	Asah1	Sp140	Mx1
Sema4d	Cmpk2	CD180	Irf2bp2
Fkbp15	Epsti1	Frmd4b	Lamp1
Dusp22	Id2	Rara	Ifitm6
P2rx7	Irf8	Itgam	Cyfip1
Lsp1	Mpeg1	Slc6a6	Slc12a9
Ctss	Gcnt1	Clec7a	Lyz2
Tfeb	Rtp4	BC147527	Tagap
Klrk1	Bri3	Plekhf2	Ighj1
I830077J02Rik	Soat1	Cd300c2	Ctsb
Mx2	Grk2	Klf13	Pira2
Gm4992	Ptprc	Gm11787	Rin2
Ms4a6c	Evl	Dusp5	Gm6169
Pstpip2	Aph1c	AbCD1	Gm9844
Ctsz	Pacs2	Rps6ka1	Rmnd5b
S100a10	Plekho1	Apba1	

Table 7: Genes downregulated during culture in Flt3L differentiation media

138 genes differentially downregulated exclusively in *CEBPA* N321D-transduced cells

Siglecf	Hist1h2ab	1810037I17Rik	Abcb1a
Fam60a	AW146154	Srxn1	Rapsn
Mmp8	Ighv1-77	Spred2	Hsd11b1
Zfp462	Hist1h2ae	Ammecr1	Pvr
Als2	Pfdn4	Pomgnt2	Zfp704
Steap3	Rgs18	Nrp2	Ifi214
P2rx3	Hist1h2an	Man1a	Ubxn11
Trim68	Six5	Nudt1	Tmem29
Ighmbp2	BC003965	BC026585	Flt1
Homez	Pon3	Depdc1b	Gm43431
Zbtb16	Zfp758	Gpr65	Kdm1a
Adgra2	Hist1h1d	Cebpzos	Hist1h3b
Rgs1	Optn	Hist1h3g	Asb1
Adcy9	4930486L24Rik	Tet1	Rps6ka3
Vcan	Suco	Polm	Tex30
Pi16	Gm37354	TlCD1	Hist1h2bg
Prcc	Fpgt	Xrcc5	Ifi47
Smtn	Ctla2b	Vav3	2810408I11Rik
Sirt1	Klhdc1	Hist1h3e	Clec5a
Slco3a1	Hist1h4n	St3gal6	Rps29
Dusp19	Actn1	Ube2w	Trem1
AA474408	Tyw3	Ankrd50	Gfpt1
Gm26597	Gfi1	Atp11b	Ift81
Csf3r	Ankrd27	Hpgd	Mettl18
9330175E14Rik	Tgfb1	Amz1	4933431K23Rik
Gtse1	Fcgr3	Erlin2	Mmgt1
Prdm1	Zfp27	Plcb2	Hivep2
Cdk5rap2	Tnni3k	Tgfbi	Tmem132d
Fam46a	Casp7	Atp10a	Zbtb34
Pam	Fosb	Il11ra1	Rims2
Gm5424	Hist1h1b	Gm7805	Anxa3
2700038G22Rik	Crim1	Pde4dip	Gm13313

Rsu1	6030408B16Rik	Nars2	CD84
Zc4h2	Rmnd1	Hivep3	
Ebna1bp2	Slc11a1	Rbakdn	

1060 genes differentially downregulated exclusively in EV-transduced cells

Selenbp1	Pum3	Tspan2	Zak
Sergef	Ldlrad4	Fancf	Kpnb1
Bhlha15	Fam114a1	Asf1a	Pced1b
Ruvbl2	Ifi27	Amot	Atr
Ifrd2	B3galnt2	Col11a2	Ninj1
Gm13502	1110051M20Rik	1110008L16Rik	Hemk1
Angptl4	Fasn	Snapc4	Cirh1a
Gm1976	Gm43213	Cdk4	Gm5593
Dennd5b	Gcsh	Fkbp3	Prkca
Zfp979	Abhd11	Cand1	Rai14
Cnpy4	Arv1	C030034L19Rik	Mroh2a
Bcl7a	Pgm2	Haus1	Pomk
Kcnq5	Cenpv	Jpx	Rpl31
Dlat	Ptgir	1110004E09Rik	Armcx6
Gm21807	Gas2	Gm14769	BC055324
Rps6	Thy1	Ccdc40	Bbs12
Ak2	Gm37510	Polr1c	Rtn4r
Zbed4	Akap7	Ppcdc	Ppfia4
Snhg4	Gm37666	2610008E11Rik	Cry2
Strbp	Polr3g	Ssbp1	Tpi-rs11
Ppfibp1	Lrrc20	Rsl1d1	Zfp335os
Jade2	Socs5	Ogt	Msantd3
Zfp566	Ktn1	Mei4	Got2
Psmg2	Pla2g4a	Ctnnal1	Gm38391
Scd2	Atat1	Cd302	Lyrm2
Cenph	Itih5	Cct3	Nck2
Brinp3	Stam	Exo5	Borcs7
Mns1	Mrpl44	Mybbp1a	Haus3
Gm5837	Gm37124	Spice1	Prep
Dennd6b	Amacr	Il31ra	Timm10
Gm27219	Dcun1d5	Chchd5	BiCD1
Gm12115	Tmem254c	Abce1	Mmp14
Rps12	Gm15659	mt-Nd2	Zfp507
Cd24a	Sort1	Bcap29	Zfp748
Zfp932	Eif3j1	Thyn1	Gm7244
Gm4879	Sh2d5	Hyal2	Pdrg1
Cep57l1	Mgmt	Paip2	Gm8459
Gm6501	Klf16	Tex9	Snora68
Ercc6	Stk16	Nip7	Cdca7l
Nutf2	Kctd4	Cisd3	Gm8173
Ctps	1810014B01Rik	Ptprcap	Zcwpw1
Gm3362	CD81	Gm8069	Gart
Mrto4	Gm12955	Fchsd2	Atp5g1
Plscr3	Dync2li1	Hsp90ab1	Zfp597
Utp20	Kdelc1	Cluh	mt-Te
Slc45a3	Trappc13	Gm9824	Pitrm1
NudCD1	2810002D19Rik	Fam78a	Utp11l
Tsen15	Dnmt3b	Ovca2	Dnajb4
Camsap1	Col19a1	Gm8741	Erh
Mphosph9	Plscr4	Aimp2	Nhsl1
Slc25a15	Nudt5	Il12a	Nme6
Hdgf	Mgarp	Blm	Acot2
Rnaseh2c	Msrb2	Srr	Smad3
Gm37201	Tnk2	Dand5	Ptpn7
Zfpm1	Ppp1r14c	Emilin1	Fam188b
Yaf2	Pdlim1	Ndc1	Nolc1
Hoxa6	Cep89	Gm15528	Mnd1-ps
Kcnj2	Kansl1l	Golgb1	Lins1
Sox4	Spint2	Rbmxl1	Ilf2
2610306M01Rik	Ccdc88c	PdCD4	Rrs1
Ddx21	Nav2	SCD1	Polr3h
Mrps31	Gmpr	Slc20a2	1190007I07Rik
Hadh	Urb2	Rps13-ps4	Paip2b
Smim24	Nsun2	Oxnad1	Gm23246
Pdpr	Dzip3	Npm3	Xpnpep3
Trip13	Bhlhb9	Sdc1	Mccc2
Spa17	Hyls1	L2hgdh	Alg2

Pbx4
Nme1
Trub2
Gm5946
Wdr75
Zfp606
Cdk6
D16Ertd472e
Sh3rf1
Gm13700
Sos2
Akap1
Ints6
Nmnat3
Dirc2
Zfp113
Tnpo1
Vangl2
Dnph1
Spata24
Gm5297
Fam212a
4930509E16Rik
2810004N23Rik
Wbscr27
Zscan22
Ntmt1
Armcx2
Hspa13
Focad
Gm20156
Plpp5
4933404O12Rik
Eef1g
Etfbkmt
Anks1
Map9
Gm17251
Gm17936
Gm6960
Mdh2
Rcl1
Taf5l
Idi1
D3Ertd751e
Slc9a3r2
4921524J17Rik
Pcbd1
Rnf220
Mfsd13a
Prtn3
Gfm1
Oaz2-ps
Paip1
Mrpl38
Slc25a29
Gm16433
Ppp1r14b
Rps2-ps10
Nup37
Rrp1b
Miga1
Pop1
Zfp146
Rpp14
Zfp783
Gm4366
Unc13a
Kctd1
Tmem186
Tspan5
Sord
Gm9847
Ncl
Mettl1
Gemin4
2900026A02Rik
Abcb6
AI987944
Tmem70
Cdca7
Prelid2
Spns2
Gm23130
Gfod2
Ube3a
Ahcy
Fam220a
Gm11993
Gm13337
Spata7
Ppargc1b
Dnajc2
Ppm1f
Chaf1b
Msh6
Ipo11
Cacybp
Gm43297
Ipp
Ccdc66
Gm37060
Osbpl5
Rptor
3110040N11Rik
Snap47
Timm9
Cep97
St14
Pcsk4
Gm1818
Aldoc
Akap13
Zfp239
Gm5864
Dusp12
Slc43a1
Rpgrip1
Rcn1
Abhd6
Gprasp1
Selenon
Fam35a
Aldh6a1
Layn
Gm15770
Zfp280c
Pde4d
Fkbp4
Aak1
Rbak
Gm13998
Bmi1
Gfi1b
Gm12848
Gm8522
Gm16465
Per3
Vwa8
Galk1
Gm14681
2610524H06Rik
Zfp329
Mthfd1
Hspd1-ps3
1810032O08Rik
Gm15654
Gm5455
Twist1
Pi4k2b
Kcnip3
Psmg1
Ska2l-ps
Prps1
Ttll1
Mex3a
Zfat
Wdr34
Mrpl15
2610307P16Rik
Polr3d
Dusp18
Nup43
Gm24452
B9d1
Ftsj3
Tom1l1
Zfp707
Mcee
Enkd1
Gm5847
Clpp
Gm42814
Ube2cbp
Gsr
Snhg6
Eif1ax
Hsd17b10
Kif17
Hist1h2bk
Pgam1-ps2
Rpl36
Ppan
Nop2
Mettl4
C230034O21Rik
Lman1
Mdn1
Sntb2
Gm6640
Nt5dc2
Comtd1
Hoxa7
4930581F22Rik
Bambi
Tceanc
Pear1
Gm4735
Xpo7
Sumf2
Qsox1
Gem
Gm9531
Nefh
Gm15261
Dhx30
Gm15800
Rpusd2
Mcm2
Susd1
Phlda3
Dnajc10
Gm30074
Gtf3c4
Bora
Pa2g4
Gm4353
Exosc2
Gm13719
Ripk3
Gimap1
Pgam1
Nop10
Hspd1
Kti12
Ppif
Cmss1
E130308A19Rik
Mcm7
Mesdc2
Unkl
Gm11451
D630045J12Rik
Nol8
Phgdh
2810403D21Rik
Tmem69
Traf4
Cbx2
Ddx1
Vsig10l
Gm4535
Gpatch4
Cngb1
Pcp4l1
Zfp40
Msi2
6330562C20Rik
Gm14857
Thsd1
Tspan13
Gng12
Hoxa5
Gm5385
Rsbn1l
Gm14057
Slc7a6
Pi15
Zcchc11
Fto
Lamp3
Klhdc2
Pabpn1
Dok4
Gm17709
Gm11914
Cd27
Usp31
Icam2
Zfp874a
Znhit3
Sgsm2
Socs2
Larp1b
Gm11516
Nop58
Tia1
Gadd45gip1
Zfp672
Spag4
Tesc
Gm11896
Las1l
Slc18a1
Gm19031
Ppp1r13b
Gm37422
Med29
Mnd1
Cerkl
R3hcc1
Hspa9-ps1

Engase
Cmtr2
Tnnt1
Uevld
Grwd1
Hmgn5
Ubxn8
Calu
Smyd2
Gm10131
C1qbp
Hddc2
Gm20342
Bsn
Gm13368
1190002N15Rik
Xk
Tnfrsf26
Shmt1
Iqsec1
Hmgb3
Cnpy2
Gstp-ps
Pex10
Ipo5
Ndufaf2
Ptcd3
Gar1
Abcd2
Mtx3
Golim4
Gpam
Mrps28
Rpl27
Gm8520
Myl4
Ctxn1
Gm11847
Lmo2
Zbtb38
Tal1
Gm17230
Slc25a43
Fmc1
Gm4737
Cct6a
Pold2
Tspan32
Top3b
Mvb12b
Pebp1
Rfc4
Zfp687
Gm7901
Zfp101
Utp14b
Gm5069
Hells
Tcp1-ps1
Snhg5
Fancd2
Rttn
Bcat1
Pecr
Irak1bp1
Npm1
Tfap4
Tmem67
Ccdc181
Rasd1
Cand2
Klhl12
1110038B12Rik
Lamc1
Prdx6
Gaa
Snord22
Mrpl50
Gm9260
Zc3hav1l
Ndufb2
Rapgef3
Ada
Polr3k
Gm10819
Gm9755
Dkc1
Tmem107
Tnfsf13os
Capn3
HaCD1
Jmjd8
G2e3
Map7d2
Tnfrsf22
Syncrip
Abcf2
Nudt19
Sirt5
Dnaja3
Tma16
Cbx3
Eno1b
Drap1
Gm10110
Ccdc167
Nudt3
Cbx5
Thap2
Dtx4
Gdf11
Clstn1
Luc7l
Wdr12
Ddx56
Ptch1
Ell2
Cdkn2aipnl
Ippk
Eef1e1
Wwc2
Pcyox1l
Gm13498
Trim28
Gm17066
Vpreb1
2610318N02Rik
Car8
Fxn
Dpy19l3
Tfdp2
Rrn3
Dtymk
Amer1
Nifk
Rmnd5a
Proser3
Txnrd3
Cycs
Mmd
Cnn3
Fabp5
Irx2
Spry4
Tnik
Gm7512
Sdhaf3
Dach1
Abhd2
Scly
Fbxo25
Etv4
Ldlrad3
Gm17837
Zfp512b
Gm5874
Lyar
Ankrd13b
Prune2
Abcb8
Taco1
Snhg15
Mif-ps3
Mtf2
Ttll5
Grhpr
Gm8113
Kremen1
Bag2
Fam216a
Rpl21
Gm42418
Rhoj
Nhp2l1
Ampd2
Slc35d3
Stk26
Hibch
Nomo1
Gm12020
Gm10020
Gm9826
Sfxn1
Gm4349
Mir5099
Rhoq
Gm16341
Dhx33
Stxbp4
Myo19
Fancb
Gsta4
Gm20696
Micall2
Fktn
Wdr5b
Rdh13
Raet1d
Elk3
Mrps18b
Eif3b
Gnpat
Dnajc12
Zdhhc24
Afap1
Hexim2
Dars
Rbp1
Paics
Gm13776
Zfp428
Serf1
Trp53bp1
Mycbp
Umps
Gm13552
Brix1
3110082I17Rik
Fam65a
Gm16379
Gm21887
Zfp775
Mrpl12
Sephs2
Wdr74
Noc2l
Gm15484
Bzw2
Orc2
2810025M15Rik
Htra3
Adsl
Cdk7
Gnl3
Gata2
Lyrm9
Tdg
Mgat5
Nob1
Repin1
Kcnh7
Frat2
Acaca
Wdr55
Ern1
Epha4
Nlgn2
Ccdc58
Gm26917
Nme4
Smpd4
Rint1
Sssca1
Gm7666
Gstp1
Fgf10
Exoc8
Wdr6
Selenom
B230369F24Rik
Gm7781
Tomm5
Fam160a2
B3gnt2
mt-Tt
D5Ertd605e
Kdm5b
Ttc3
Mrgpre
Gm9256
4732463B04Rik
Rrp12
Gspt1
Mcph1
Guf1
Ruvbl1
Slc16a1
Poli
Prmt5
Trp53rkb
Ahi1
BC017158
Gm13937
Syce2
9930104L06Rik
Tufm
Ccnh
Asap2
Gm13521
Tmem97
Tarbp1
Gm807

Ankrd26
Gclm
Tfpi
Rpusd3
Trnt1
Ttc37
Gm6166
Polr2f
Pam16
Aldh18a1
Hgh1
Gm5525
Pom121
Rpl36-ps2
Wdr77
Stxbp5
Ddah2
Ints3
Mkrn2
Idh3a
Hsd17b14
Fam92a
Rps8
Zfp369
Nop56
Hcst
C920006O11Rik
SpoCD1
Nup133
D130058E05Rik
Snord55
Brwd3
Raet1e
2610301B20Rik
Gm5611
Tfb2m
Tctn3
Psmd7
Rfx3
Zfand4
Mthfd1l
Dnajc27
Idh2
Cad
Suv39h2
Katnal1
Adgrg3
Fubp1
Fam133b
Gm9797
Eno1
Ahdc1

Pde8a
2410016O06Rik
H2-Q6
Bola1
Srek1ip1
Dut
Ddx18
Cd99l2
Gm14109
Jmjd4
Aasdh
Rpl36a
Wdr36
2610002M06Rik
Pcyt2
Osbpl7
Gm5687
Ypel1
Slc29a2
Ap4e1
Lsmem1
Hist1h2bj
Myb
Cep41
Gm4258
4930520O04Rik
1700030K09Rik
Mcm6
Yod1
Snx27
Phc1
Rai1
1810022K09Rik
Pus7l
Chtf8
Dph5
Ppat
Mmachc
Gm26580
Dnm1l
Rad18
Bace1
Ino80dos
Slc37a1
Nipsnap1
AI506816
Itgb3bp
Ift27
Pomgnt1
Zc3h7b
Nt5c3b
Mterf2

Camkk1
Fkbpl
Prdx3
Zfp184
Arap2
Faf1
Fer
Qtrtd1
Ccdc15
Gm17494
Phb
Pofut1
Pus7
Smyd5
Mc5r
Cd72
Gm5855
Zbed3
Gm7556
Simc1
Npm3-ps1
Nudcd2
Ubfd1
Pif1
Adat1
Atad3a
Gemin5
Gm7332
Pgk1
Mlh3
Rcc1
Mettl15
Chordc1
Rnf219
Gm7289
Msto1
Zfp799
Pla2g6
Hspbp1
Hist1h2ac
Gm6652
Mzb1
Gm6505
Zfp503
Gm5703
Rasa1
Gm5251
Prdx6b
Oaz2
Tpi1
Atg12
Etfb

Gm5124
Spout1
Ccdc125
Klhl9
Amd1
Senp8
Pabpc4
9230114K14Rik
Gm14328
Rpl12
Cbx6
Dusp4
Lekr1
Ttc26
Gm6447
Psme3
Pus10
Xpc
Nudt7
Kcnq1ot1
Arid4b
Srm
Yae1d1
4930555A03Rik
Mlec
Cbr1
Thop1
Mcm3ap
Gm7251
Maged2
Polr1b
Gm12583
Gm16074
Pitpnc1
Gimap6
Flt3
Pdss2
P3h1
Nlrx1
Gm10384
Gm7933
Dctd
Taf1a
Fbxo9
Plppr3
H2-Q7
Vkorc1l1
Zfp219
Crnde
Trap1
Arpp21
Zfp160

280 genes differentially downregulated in both *CEBPA* N321D and EV-transduced cells

Eya1
Car3
Ybx3
Mmp11
Tha1
Prkar2b
Gm12611
Al314180
Ift57
Samsn1
Gnpnat1
Hist1h3f
Ttc39c
Chchd10
Epha7
Meis1
Ydjc
Kit
Plod2
F2rl3
Pfkl
Sipa1l1
Tst
Rcn2
Hoxa10
Olfml2b
Sh3d19
Itga6
Zfp703
Gse1
Por
Cela1
Pnkd
3000002C10Rik
Sigirr
Gjb3
Rasgrp2
4930519L02Rik
Gcnt4
Clec4e
Ocrl
Reps1
Kcnq3
Calcrl
Hist1h1a
Nrg1
Dab2ip
Gab1
Ascl2
Flnb
Limch1
Prodh
Angpt1
Gca
2300009A05Rik
Hist1h3d
Slc17a9
Nicn1
Gfra1
Hmgcr
Mt1
Fat4
Gimap9
1700012B09Rik
Ppid
Cited1
Adgrl4
Dok2
Asph
Ppp1cb

Ndufa4
Ebf1
Papss2
Tfrc
Ccl9
Gm16194
Gm9143
Mab21l1
Ube2o
Dab1
Clec4d
Spry2
Col4a5
Mtap
Rnf125
Aebp1
Prss57
Ssfa2
Erg
1700113H08Rik
Gm10169
Mt2
Rbpms2
Chsy1
Ddx4
Gm43096
Endod1
Acot3
Bphl
Psmd5
Bmx
Plscr1
Gm29397
Hilpda
Gtf2f2
Gm17149
Gm11349
Tespa1
Gm16897
Hlf
Atp11a
Igkv12-89
Bhlhe41
Igll1
Il6
Fam162a
G6pc3
Impdh1
Dlg2
Vat1
Anxa9
Gimap5
Acot1
Dtd1
Galc
Srgn
Chchd4
Nipsnap3b
Lima1
Col23a1
Slc22a4
Ltbp3
Stap1
Gata3
Efna5
Serpinb1a
1700003F12Rik
Ica1
Nrgn
Vamp5

Crtap
Hmgcll1
Igfbp7
Arhgap6
Mpo
Tsc22d1
Tspan6
9330159M07Rik
Ftsj1
Ubash3a
Ctsg
Plcb4
Ms4a3
Gm14450
Gm12403
Pcdh7
Lsm3
Mmp13
Nedd4
Slc38a5
Fkbp11
Ap3s1-ps2
Ctla2a
Lbp
Ighv1-23
Chl1
Cul7
Tacstd2
Pglyrp1
Chek1
Fam136a
Gfra2
Slc35b4
Hoxa9
Arhgap18
9630013D21Rik
Ergic1
Ppic
Cd320
Pfas
Tyms
Gm37818
Mcpt8
Gm6104
Sapcd2
Rag1
Cd34
Dntt
Lmo4
Rsph9
Armcx1
Abhd17c
Hk3
Mcfd2
Tmem40
Nfe2
Acyp1
Parp16
Ptk7
Gda
B230118H07Rik
Hgf
Ptprm
Ero1l
Gm11998
Gm14630
Col4a6
Aasdhppt
Rab44
Gm5530

Gm12158
Adgrg1
Tmem108
Dtnbp1
Gm3470
Ptger2
Csgalnact1
Itga9
Gpr12
Nrxn1
Sptbn1
Hmga2
Slc28a2
Afap1l1
Mif
Gm15658
Gm6065
Ap3s1-ps1
Sema3a
Ablim1
Plxna3
Wdr60
Gm16302
Gm5551
Gm43489
Stat4
Egln3
Dgat2
Zxdc
Gm7389
Trem3
Pde2a
Galnt14
Emid1
Mmp12
Fndc3b
Slc19a1
Elane
Gm4705
2310075C17Rik
Gm15581
Timm8a1
Cdkn2c
8430408G22Rik
Psd3
Rgcc
Dnajc6
Ttll9
Gm16104
Muc13
Ptgds
Igkv4-90
Mllt3
Nkg7
Gstt2
Rbpms
Adgrg7
Cdc14b
Psat1
Ap3s1
Zfp934
Ednra
Car2
Zhx2
Cit
St6galnac3
Armcx4
Rag2
Dus4l
Gpc3

Table 8: Genes switched on and switched off by the *CEBPA* N321D mutation

385 genes switched off by the *CEBPA* N321D mutation

Slc7a8	Gm15448	Rab32	Gm15753
Spib	Fcgr1	Atp2b1	Card19
Gbp2	Fcgr3	Gpr160	Wipf1
Fam46c	H2-Ab1	Gm7676	Lamp2
Trem2	Rsad2	Slc35f6	Morc3
Acvrl1	Cass4	HaCD4	Gm12166
Ms4a7	Arhgap30	Lpcat4	Sec24b
Fcgr4	Casp4	Zfyve26	Npc2
Ifit1bl1	L1cam	B4galt5	Efhd2
Smpdl3b	Isg20	Mcl1	Actg-ps1
Cxcl10	Cnr2	Vwa5a	Msrb1
Slfn1	Icosl	Gpr137b-ps	Ccdc86
Cxcl16	Gm5431	Peli1	Gm12164
Adgre4	Adora3	Samd9l	CD48
Fgl2	Naip6	Gm8902	Dpysl2
Ifit1	Apobec1	Irf1	Gm23374
Apoc2	Tnf	S100a11	Kctd12
Phf11d	Lmln	Nmi	Atp6v1b2
Slamf9	Adgre1	Ccnd1	Ms4a6c
Plxdc1	Gadd45g	Jak2	Sp110
H2-Aa	Dock5	H2-K2	Rgs2
Nfatc2	Mgl2	Herc6	Gm8399
Tlr13	Slc15a3	B3gnt8	Arrb2
Gm12250	Naaa	Cdk13	Tmem9b
Rgs1	Oas1a	Adam15	Esyt1
Phf11a	Ids	Hgsnat	Pik3cd
Tlr7	Lilra6	C130026I21Rik	Gm4760
C3ar1	Hcar2	Pik3ap1	Gngt2
H2-Eb1	CD40	Slc29a3	Rnf34
Il18	Tbc1d8	Eif2ak2	Snx2
AW112010	Clec4a1	Nfkbid	Cyb561a3
Aim1	Lacc1	Sema4a	Rnf114
C1qb	Sorl1	Cotl1	Flna
Fam129a	Ciart	Myo1c	Ftl1-ps1
Pilrb2	Dab2	Vasp	Nampt
Trim30c	Ifi203-ps	Plin2	Anxa2
Slc37a2	Sertad1	Sft2d1	Hexa
CD80	Hck	Crip1	Acta2
Cd22	Oasl1	Nfkb2	Syngr2
Gm37347	Ifi204	Gm7665	Cndp2
Zfp467	Plaur	Prkch	Dck
Arl5c	Dusp1	Taf13	Cstb
Fgr	P2rx4	Klf6	Tmem219
Id3	Parp12	Akt3	Acot10
Siglec1	Irak2	Dennd1b	Gm12715
Tlr4	Atf3	Gm11189	Ppp1r18
Nxpe4	Zfp36	Slfn8	Taldo1
Msr1	Anxa1	Gltp	Sh3glb1
Cxcl2	Ifi47	AC132444.1	Rhog
Bco2	Fry	Camk1	Ftl1
Maf	Tmem176b	H2-DMa	Swap70
Clec4b1	Trim5	Lipa	Gm12791
Trim30b	Zbp1	Rap2c	Rel
Ifit3b	Sirpa	Mvp	Twf2
Klk1b11	Cyp4f18	Ifi35	Gm12428
Ltb4r1	Gm15922	Itpr1	Actg1
Ccdc109b	Cttnbp2nl	Ero1lb	Ywhah
Tnip3	Gm6377	Pstpip1	Cnn2
9430076C15Rik	Il2rg	Pnpla7	Cyba
Clec4a4	Csf2ra	Tmsb4x	Rab8a
Lpl	Il15	Selplg	Ptpn1
C1qc	Tmem86a	Ncoa3	Rnpep
Dpp4	Fos	Dtx3l	Gm5548
Clec4n	Bhlhe40	Dusp3	Ddit3
Tlr1	Lrrc25	Ep300	Gm17087
Phf11b	H2-DMb1	Picalm	Scand1
Cd274	Rnf213	Gm4918	0610012G03Rik
4930430E12Rik	Il12rb2	Rcbtb2	Mapk14

Rasgef1b
Tnfsf8
Slamf7
P2ry12
1600010M07Rik
Ifit2
St18
Card11
Ifit3
Tmem106a
Oas1g
Kynu
Tgfbi
Ifi213
Hpgds
Sirpb1b
Alox5
Cdk14
Gm20559
Clec9a
Gbp4
Lyz1
St8sia1
Ms4a4a
Lrp1
Ccrl2
Aoah
Upb1
Aif1

Pid1
Grasp
Etv3
Gm6175
Gsap
Tmem150b
Junb
Fam26f
Ctsc
Skil
Ptafr
Zfp36l1
Zc3h12a
Slfn9
Egr1
Irgm1
Plekhm3
Abcg1
Fbxw17
S100a6
Ptms
Gm20432
Pfkfb4
Txndc16
Ier5
Emilin2
Pim1
Ifi30
Gm1966

Ggh
M6pr-ps
Plxnc1
M6pr
Mycbp2
Tnfrsf1a
Creb3
Gm21738
Abhd12
Hexb
Cd9-ps
Acss1
Ifngr2
Fam69a
Gm16026
Ly9
Tnfaip8l2
Stard3nl
Nrp1
Prdx5
Ppp2r5a
Gm10800
Cd9
Hist3h2a
Ifi214
Cln3
Mndal
Vim
Stk38

Ctsd
Gm8185
Actb
Ccdc50
Crlf2
Laptm4a
Iqgap1
Ttc7
Gm14856
Degs1
Necap2
Gm12543
Arf5
Gm4750
Gnaz
Gm8464
Gnai2
Wdr1
Atp6v0c
Psmb10
Cst12
Rap1b
Malat1
Gm8822
Itm2b
Ier2

133 genes switched on by the *CEBPA* N321D mutation

Zfp329
Zfp184
Tnfrsf26
Cbx2
Cd27
Abcd2
Osbpl5
Aasdh
Abhd6
Rcn1
Col19a1
Hemk1
Cbx6
Htra3
Taco1
Katnal1
Ccdc66
Afap1
4930520O04Rik
Ctxn1
Ccdc15
Spice1
Ankrd26
Pi15
Zfp160
Mgmt
Rpusd2
Bhlhb9
Oxnad1
Gm10384
Snapc4
Socs2
Amer1
Gas2

Fktn
Thap2
Pex10
Gm20696
Fabp5
Tnfrsf22
Gm1976
Gm9847
Spata7
2810403D21Rik
Gm13998
Gm14328
9230114K14Rik
Klf16
Gm6166
Zfp932
Tspan2
Zfp113
Rrp12
Tspan32
Mnd1
Mnd1-ps
Il12a
AI987944
Myb
Kif17
Rdh13
Cnn3
Cep97
Gm20342
Trp53rkb
Jpx
Yod1
Gm17936

Zfp369
Myl4
Ldlrad4
Ino80dos
Ada
CD81
Cmss1
Tspan13
Arap2
1190007I07Rik
H2-Q7
Sdc1
Hmgb3
Bag2
Dtx4
Zbed4
Elk3
Ptprcap
Gm7901
Mycbp
Phgdh
Gm5847
Xpc
Smpd4
Cdca7
Zfp687
3110040N11Rik
Bace1
B3gnt2
H2-Q6
Ifi27
Serf1
Gm5385
Adgrg3

Srek1ip1
Cdk6
Ppp1r14b
Gm5124
Dkc1
Bzw2
Ddx1
Lyar
1110038B12Rik
Gm17230
Mettl1
Gm16379
Gm13719
Snhg15
Gm7332
Gm4737
Gm15659
Gm9826
Gm10020
Ncl
Nudcd2
Paip2b
Tufm
Gm12020
mt-Te
Ahcy
Ssbp1
Gm17837
Hspd1
Ak2
Hspd1-ps3

Table 9: Genes upregulated during culture for 5 days in Flt3L differentiation media

637 genes upregulated in main population of 78 *CEBPA* N321D-transduced cells

Siglech	Il6st	Ifnar2	Gpsm3
Ms4a4c	Gm42641	Prkx	Klf13
Tpm2	Parp14	Gatsl2	Tnfrsf13b
Klrk1	Gas7	Gm14699	Gm4875
Clec7a	Id2	Hes6	Uvrag
Oasl2	Vsir	Tapbp	Trim25
Rnase6	Tmcc3	H2-T23	Bloc1s1
Oas2	Mx2	Naga	Pink1
Lair1	Ighj2	Tyrobp	Nfkbia
Cd300lf	Fads3	Gm17791	Fli1
AI839979	Trim34a	Avl9	Dleu2
Pilra	Znfx1	Gm8995	Smim3
D930048N14Rik	Lhfp	Plec	Rapgef1
Tmem51	Fyb	Bmyc	Fmnl1
Fabp4	Gm12892	D1Ertd622e	Myo1f
Mir22hg	Bach2	Necap1	Capg
Chrng	Ckb	Epn1	Dap
Kctd14	Camk1d	Fam174a	Mpc2
Smim5	Galnt10	Cnp	Satb1
Cx3cr1	Tdrd7	Trafd1	Pecam1
Slfn2	Gm15417	Tor3a	Tiparp
Myog	Gpr141	Egfl7	Fkbp15
Ly6c2	Ly6a	Iffo1	Xiap
Klhl30	Cd300c2	Gpcpd1	Ssh2
Il10ra	Rnf150	Adgre5	Sesn1
Lifr	Slc46a3	Scpep1	Gm13669
Frmd4b	Plekha1	Vps37b	Aagab
Itgax	Zfp933	Anxa6	Gm8752
Ptpro	Myadm	Arhgap9	Gng2-ps1
Clec10a	Pacs2	Stat1	Mtmr14
Dpep2	9930111J21Rik1	Lamp1	Hmha1
Alcam	Isg15	Serpinf1	Mtpn
Plekhg5	Ms4a6b	Ypel5	Prkcd
Ifi207	Gm15987	Atp1b3	Ctsz
Lcp2	Ccdc180	CD82	Alox5ap
Ahnak	Dusp7	Gm11787	Ifi35
Slc37a3	Tmem229b	Lst1	Ythdf3
Ly86	Rgs10	Spi1	Notch2
Gpr34	BE692007	Gm1848	Alyref2
Cd74	Zfp947	Gm26798	Man2b1
Lsp1	Cd28	Il4ra	Ptpn6
Ccr1	Sgcb	Mkrn1	Wdr26
Epb41l4b	Mroh1	Plac8	Runx1
I830127L07Rik	CD163	Klhl6	Nptn
Gpr183	Mpeg1	Sp100	Mef2c
Ctsh	Neat1	Gm3788	Trappc5
Csf1r	Lbh	Tulp4	Heg1
Cds1	Ctns	Pip4k2a	Tpd52-ps
Ifi27l2a	Slc12a9	Bcl2l11	Sh3bp1
Sgcg	Ap1s2	Tubb6	Hcls1
Il1r2	Clec4a3	Bcl2	Gosr2
Cd300c	Ypel3	Hps3	Mef2d
Pirb	Ube2l6	Ctsb	Prr13
Lgals3	Parp11	Fam105a	D930015E06Rik
Dusp22	P2rx7	Plxnb2	Zeb2
Nlrp3	Fkbp1b	S100a4	Pitpna
Klrd1	N4bp3	Ocel1	Fam168b
Rbm47	D8Ertd82e	Zbtb20	Gm28875
Sema6d	Irf9	Fam65b	Man1b1
Pld4	Rasa4	Soat1	Nfatc1
Clec4a2	Tmed8	Mef2a	Sppl2a
Gcnt2	Evi2a	Hoxb8	Plekhf2
Oas3	Borcs6	Bst2	Tm6sf1
Csf2rb2	Filip1l	Epsti1	Mpc1-ps
Tpst1	Fam234a	Anxa5	Ivns1abp
March1	Hotairm1	Gm7285	Gm6169
Itgb7	Tlr2	Mkrn1-ps1	Gm12854
P2ry13	Itgb2	Pak1	5730416F02Rik
Usp18	Themis2	1700025G04Rik	Cirbp

Apba1
Tnfrsf23
Tnfsf9
Slc14a1
Mx1
Ctnnd2
Ctss
Tagap
Rin2
Ighj1
Rubcnl
Slc9a7
Slc16a7
Six1
Il13ra1
Lgmn
Gm7694
Cysltr1
Kmo
Scin
Abcb4
Tns1
Klrb1c
Cfh
Pstpip2
Clec2i
Pira2
Lilrb4a
Ctse
Itgam
Tfeb
BC147527
Tmem8c
Cmah
Nos1ap
Sdc3
P2ry6
Lilr4b
Slfn5
Ighm
Ms4a6d
Mylpf
Wfdc17
Klra1
Cybb
Irf7
Gm14548
Ifi211
Rasgrp4
I830077J02Rik
Ifitm6
Cox6a2
Mycl
Arl4c
Bmf
Klrb1f
Hpse
Fam214a
Aph1c
Ak8
Havcr2
Kdm7a
Evl
Tnni2
Klf3
Gm15931
Rtp4
Dhx58
Gsn
Map3k8
Pik3r5
Ifi205
Rogdi
Cd68
Sik1
Prr5
Cd300lb
Psap
Slc44a1
Grn
Gpr137b-ps
Nceh1
Ms4a4b
Osbpl3
Atp7a
Tubb2a
Abhd15
Rab7b
Cyth4
Plekho1
St3gal5
Tecpr1
5031439G07Rik
H2-T-ps
Notch1
Stat2
Inpp4a
Dcaf10
Unc93b1
Gm5398
Tep1
Rcan3
Tmem229b-ps
Lpxn
Tmem119
Blvrb
Selenop
9930111J21Rik2
Sdc4
Sema4d
A530040E14Rik
Gm5552
Cst3
Dusp5
Zbtb18
E330020D12Rik
Sfxn3
Aldh3b1
Usp12
Cd52
Gab3
2310001H17Rik
Irf5
Cmpk2
Cd300a
Gfod1
CD14
Inpp5k
Plcl2
Mtm1
Fam134b
Gm14608
Acox3
Xaf1
Trpv2
Btg1
Trim12a
Fgd2
Map3k5
Trim30a
Sat1
Pigv
Il6ra
Sh2b2
Pmaip1
Tubgcp5
CD180
Fbxw4
Irf2bp2
Ifi203
Fads2
Sh2b3
Prex1
Gm15989
Stk17b
Zfp296
Sft2d2
Cytip
Gm14032
Pafah1b3
Dr1
Zufsp
Parp9
Gm4992
St8sia4
Lgals3bp
Zfp263
Pisd-ps1
Tifa
Arhgap17
Arhgef12
Bex6
Ppm1m
Lyz2
Rprd2
Hlx
Zfp62
Gns
Dhrs3
Map4
Tspo
Snx20
Mbnl1
Lyn
Fam76b
Sp140
Rpph1
Fam129b
Tmem55b
Rmnd5b
Ehd4
Kcnk12
Nfam1
Tmem109
Atp13a2
Itm2c
Lnpep
Map3k1
Slc2a6
Cnbd2
Ctnna1
Zbtb7a
Milr1
Gm6088
Runx2
Tec
Laptm5
Cebpd
Cdkn1a
Gm2622
Napsa
Gm9844
AbCD1
Smim14
Ccnd1
Selenow
Rsrp1
Txnip
Rab5a
Smim20
Ctsa
Gm12345
Syk
Ccl3
Parp8
Abi1
Lasp1
Tmed3
Gm8181
S100a10
Trpc4ap
Gm8666
Akirin1-ps
Was
Ogfr
Cyfip1
2010001K21Rik
Whsc1l1
Ifitm3
Plp2
Asah1
Rps6ka1
Arl6ip5
Znrf1
Bnip3l
Usf1
C330007P06Rik
Acaa1a
Fndc3a
Ubl3
Gm11663
Slc44a2
Gm17887
Cmtm6
Grk2
Vamp2
Ddx24
Litaf
Nrros
H2-T22
Rassf5
Rnf130
2610206C17Rik
Tor1aip1
Nup210
Arpc5l
Fes
Slc6a6
Msn
Ap3b1
Ncoa4
Pcyt1a
Bloc1s2-ps
Nkiras2
Luc7l2
Gm16372
1110059G10Rik
Setd2
Wasf2
Dgkd
Pycard
Serinc3
Zfp36l2
Reep5
Ptprc
Leng8
Limd2
Tifab
Sympk
Cdc42se1
Bin1
Sh3bgrl3

Itga1
Bcl6
Naip5
Irf8
Snx29
Cln8
Slc8a1
Rara
Ccl6
Ighj4
Rassf4
Axl
Srpk3
Ebi3
Ms4a6c
Arhgap5
Trim30d
Shisa2
Cadm1
Apc
Gm8615
Csf2rb
Gng2
Samhd1
Med13l
Slfn8
Tmem170b
Zfp385a
Gcnt1
Uba7
Ly6e
Shisa5
H1f0
Stx7
Cfp
Slc27a1
Igfbp4
Gm12045
Tet2
H2-M3
Ube2a
Tpd52
Gm7936
Acer3
Atp6v1d
Il17ra
Tmsb10
Tmem50b
Bri3
Cbfa2t3
Snx30
Cd93
Adcy7
Tcirg1
Gabarap
Pcif1
Iqgap1
Arhgef6
Fxyd5
Hmgn1
Zc3hav1
Gm7823
Edem1
Tmem261
Gm12420
Dpysl2
Cdc42se2
Cyb5r4
Flna

21 genes upregulated in subpopulation of 5 *CEBPA* N321D-transduced cells

Ly86
Csf1r
Ccl3
Plac8
Cd52
Lst1
Gm2965
Fcer1g
Rps15a-ps7
Gm14539
Ly6e
Atp6v0e
Mpc1-ps
Dynlt1b
Gm21399
Rps15a-ps5
Gm8430
Prdx1
Gm15148
Rpl30
Lgals1

1399 genes upregulated in EV-transduced cells

Ly6a
Mx1
Axl
Ms4a4a
Ifit1
Trem2
Spib
Slamf7
Tmem51
Frmd4b
Ifit3b
Slfn1
Phf11d
Cysltr1
Tnfsf8
Klra17
Acvrl1
Plxdc1
Tlr1
Gm12250
Itgax
St8sia1
Ifit1bl1
Nfatc2
Scimp
Rras2
Epb41l4b
Apoc2
Mycl
Gbp9
Tlr9
Ifit3
Id3
Pid1
Oas2
Il13ra1
H2-Aa
Siglech
Cd74
Fgd3
Gsn
Rasa4
Tmem86a
Naga
Cyp4f18
Sat1
Shisa5
Ddx58
2810405F17Rik
Zfp467
Atf3
Nfkbid
9430034N14Rik
Klf3
Epsti1
Slfn8
Soat1
Sdc4
Rogdi
Isg20
CD14
Fkbp1b
Susd3
Tlr4
Il12rb2
Crybg3
Irak2
Sh3tc1
Cd68
Gm5552
Zfp36
Zbtb7b
Trim34a
Lmln
Cln8
Ier5
Hpgds
Zfp382
Trim12c
Bcl6
Ly9
Ifi209
Nckipsd
Sema4d
Tifa
Daxx
Gm4918
Pdlim5
Cebpd
Tmem9b
Gpr132
Pmaip1
Spi1
Lyn
Mgat4a
Hist3h2a
Fmnl1
Lpp
Gm8995
Nfkbiz
Slc15a4
Tspo
Arrb2
Ypel5
Rab32
Ptger4
Eif2ak2
Arhgap17
Diaph2
Sik1
Cotl1
Zc3hav1
Sppl2a
Anxa4
Il15
Tnni2
Creb3
Trim25
Tbc1d13
Ppp2r5a
Osbpl9
Atg7
Pik3ap1
Hk2
Coro1a
Grina
Tle3
Abcg1
Bcl10
Med10
Arhgap25
Slc25a45
Tet2
Ubl3
Pitpna
N4bp2l1
Tubgcp5
Rap2b
S100a6
Nampt
Gpcpd1
Aff1
Acss1
Foxred2
A230046K03Rik
Usf1
Ubc
Capn2
Gm13212
Actg-ps1
Sdccag3
Sbno2
Pisd
Ptk2b
Sirt7
Fuca1

Oasl1
Lilra6
Ms4a4c
Adgre4
Lilr4b
Clec4a3
Ctnnd2
Trim30b
Dpep2
Gbp4
Fcgr1
Ms4a7
Bcl3
Slc37a3
Ctsh
Ccdc109b
P2ry13
Fam129a
Tlr7
Ahnak
Ly86
Pik3r5
Slfn2
Rsad2
P2ry10
Oasl2
Ifi207
Cybb
Ly6c2
Csf1r
Lilrb4a
P2rx7
Clec7a
Sdc3
Ifi205
Ptpro
Clec4a1
Cxcl16
Klrk1
Phf11b
Slfn5
L1cam
P2ry6
Cfh
Usp18
Ccr1
Zbp1
Ifi27l2a
Ifi211
Cd300lf
Clec4n
Tpst1
Aif1
Rnase6
Pilra
Lgals3
Tlr13
March1
Irf7
Cttnbp2nl
Cass4
Fads3
Slc15a3
Rin2
Nlrp3
Clec4a2
Gpr34
Lsp1
Trim30d
Fcgr4
Gm14548
Msr1
Klrd1
Vwa5a
Rara
Cnr2
Prkch
Lrp1
Arhgap30
Pstpip2
Ighj2
Lrrc75a
Ctsb
H2-DMb2
Metrnl
Ebi3
Adap2os
Gm14699
Lpin1
Rnd3
Filip1l
Tnfaip2
Slc9a7
Raph1
Fbxw17
Fry
Il6st
Gm15503
Gas7
Themis2
Slc12a9
Ccrl2
Ly6c1
Neurl3
Plpp6
Cd300a
Itga1
Pak1
Naip5
9930111J21Rik2
Ccl4
Slc16a7
Rasgef1b
Inpp4a
Gm3788
Junb
Lpxn
Cst3
Gm14032
H2-T23
Aph1c
Sirpa
Arhgap5
Nfam1
St3gal2
Maml1
Ccdc180
Skil
Rgs3
Herc6
5031439G07Rik
Gm9844
Fgd2
Snx21
Armc7
Grasp
Klf4
Trafd1
Zcchc2
Slfn9
Zfp36l1
Tubb2a
Slc29a3
Tmsb10
Ap1s2
Lbh
Nfkb2
Prkcd
Tax1bp3
Sft2d1
Slc43a2
Iqgap1
Itpripl1
Mcl1
Efhd2
Tnfrsf1a
Ppm1h
Lipe
Ncoa3
Tcf19
Gm7665
Tcirg1
Arid5a
Gm7357
Gm11787
Akna
Lpar6
Npc2
Gm4875
Net1
Lasp1
AC133103.1
Ptpn6
Dennd1c
Slc25a22
Ctsa
Gm15753
Itm2c
Dck
Gm16205
Fxyd5
Dtx3l
Chmp4b
Nkiras2
Litaf
Map4
Tnfaip8l2
Zyx
Cmtm3
Itpr1
Pag1
Acbd5
AbCD1
Rgs2
Asah1
Gm21975
Mpc1-ps
Plekhg3
Laptm5
Plec
Lgals8
Syngr2
Fbrsl1
Dab2
Sh3glb1
Cdkn1a
Cmtr1
Pim1
Nmi
Snx2
Taf13
Gm4992
Peli1
Epn1
Nedd9
Tmem229b
H2-T22
Wdr44
Spaca9
Ccl3
Rnf167
Ftl1
Gm9025
Ifih1
Wipf1
Ostf1
Vrk1
Gm3145
Cib1
Sap30
Map3k1
Il3ra
Fcgr3
Atp6v0c
Zfp277
Rnf34
Ifngr1
Mob1a
Pttg1
Dbnl
Myo1c
Msrb1
Cd244
Gm6615
Sh3bgrl3
BC005537
Gm10801
Fcer1g
Bin1
Arpc5
Ctnna1
Hcls1
Gna13
Nfe2l2
Sgk3
Limd2
Mfsd5
Hmha1
Gm7823
Psmb10
Elmo2
Rbpj
Tmed3
Cd2ap
1600014C10Rik
Cd53
Ncf2
Sirt2
Akr1a1
0610012G03Rik
Gm7936
Cln3
Whsc1l1
Prkd3
Selenow
Cbl
Ccdc86
Rchy1
Plin2
Sec24b
Racgap1
Erlin1
Gabarap
M6pr-ps
2310035C23Rik
Lrrk1
Ppp1r18
Cst12
Cdc42se1
Stat6
Rraga
Calm1

Cd300c
CD40
Il10ra
Rasgrp4
Isg15
Gm20559
H2-Eb1
Ifitm6
Naaa
Cx3cr1
Dusp22
Tnip3
Mx2
Gm5431
Gbp2
Apba1
Aim1
Oas1a
Gpr183
Cxcl10
Itgb7
Ctss
Lgmn
Pirb
Smpdl3b
Fyb
Slc37a2
Dock4
C3ar1
Phf11a
Pilrb1
Pira2
Clec10a
Il1r2
Btg2
C1qb
Cds1
Tnf
Kmo
Pilrb2
Trim30c
AI839979
Fgl2
Hcar2
Ccl2
Itgam
Pld4
Oas1g
Il18
Id2
Cd22
Ms4a6d
Card11
Tnfsf9
Lyz2
Arl5c
Stard9
Ccl6
Atp1b1
Ifit2
Ifi204
3110043O21Rik
Cd300c2
Gadd45g
Ctse
Chn2
Tmem154
Gpr137b-ps
Fgr
Ube2l6
Upb1
St18
Parp14
Camk1d
Cd300lb
S100a4
Mdfic
Scpep1
Zc3h12a
Plaur
Sp140
Gpr146
Ifitm3
Rnf213
Havcr2
Casp4
Lrrk2
Dusp1
Galnt10
Fcor
Rell1
F630028O10Rik
Arhgef12
Il4ra
Plekho1
Gsap
AW112010
Tecpr1
Cadm1
Btg1
Ifi35
Csf2ra
Ltb4r1
Plekha1
Dusp5
Prr5
Sp100
Etv3
Bst2
Trpv2
Igsf6
Jak2
Amn1
Mir22hg
Lacc1
Tlr2
Ccnd1
Tnfrsf14
Neat1
Myo1f
Anxa5
Gm8902
Ifi203
Sorl1
Slc46a3
Gm43330
Cfp
Ncoa7
Plxnb2
Gm15417
Tdrd7
Slc35f6
Bcl2l11
Padi2
Frmd4a
HaCD4
Itpripl2
Paqr4
Tapbpl
Carmil1
Lyst
Gm37589
Nr4a1
Plxnc1
Prkab2
Ly6e
Dnajb14
Plac8
Dock5
Nuak2
Gm20432
Cyp4f16
Necap1
Cmtm6
Ankfy1
CAAA01147332.1
Pcsk7
Mycbp2
Birc3
Gm13392
Gm7676
Ptpn1
Sertad1
Myd88
Mbnl1
Vim
Usf2
Dennd1b
Dram2
Tulp4
Tmem184b
Degs1
Zfp992
Gng2-ps1
Slc25a38
Vps37a
Rnf115
Gm43245
Ptpn22
Sp110
Gm5578
Rps6ka5
Milr1
Cyba
Antxr2
Pip4k2a
Adam15
Lrch3
Man2b1
Gpr171
Fam129b
Gm4760
1110008P14Rik
Mov10
Tfe3
Stard3nl
Ap2a2
Gng2
Cstb
Arhgap9
Klhl6
5730416F02Rik
Rcbtb2
Gm5046
Cdk13
Stard3
H1f0
Gm13535
Rnf114
Stxbp3-ps
Sgpp1
Rap2c
Lmo1
Prkx
B4galt5
Capg
Brox
Rnf166
Gm12003
Gpr108
Gng10
Leprot
Rpph1
Gm6263
Hmgcl
Nrros
Slc35c2
Aim2
Vasp
Inpp5k
Fnip1
Arf5
Gm17791
Hpcal1
Lactb
Actg1
Pepd
Tnfrsf13b
Cnn2
Nt5c3
Fnbp1
Hcfc1r1
Ccdc88a
Slc34a2
Polb
AB124611
Esyt1
Dcaf11
Gm6977
Ttyh3
Cmip
Gm5590
Slc44a2
Wdr45b
Lcp1
Rp2
Hexa
Zbtb7a
Gm8577
Gm8822
Tmem173
Ifnar1
Tm2d2
Gm15487
Max
Map7d1
Mical1
Tmem243
M6pr
Fkbp15
2010001K21Rik
Anxa2
Epb41l2
Renbp
Slc35c1
Acaa1b
Lrrfip2
Acot10
Prr13
Ostm1
Ascc3
Acap2
Actr3
Fth1
Nipa2
Gm12715
Gm6305
Psme2b
Gm14856
Spop
Sp2
Tet3

Rbm47
Tmem106a
0610040J01Rik
Evl
Tfeb
Clec4b1
Oas3
Adgre1
Fam46c
Dram1
Maf
Lifr
Gm37347
Ifi208
Gpr18
Adora3
Siglec1
Smim5
Gm15987
Ms4a6b
Adap2
Kctd14
Nxpe4
9430076C15Rik
C3
Ifi206
P2ry12
Gm15931
Alox5
Wfdc17
Cd300lg
Socs1
B430306N03Rik
Naip6
Csf2rb2
Klk1b11
Cxcl2
Slc14a1
Grap
Cmpk2
Bco2
I830127L07Rik
Ms4a4b
Lcp2
Bach2
Lyz1
Clec4a4
Znfx1
Gm15922
C1qc
Gm15448
Vsir
Socs3
Tbc1d8
Gm5398
Dpp4
Parp12
Dusp10
Tagap
Lpl
Fcgrt
Lair1
Ighj1
Kynu
Cd274
Myadm
Irgm2
Sirpb1b
9930111J21Rik1
Ifi213
4930430E12Rik
Zfp385a
Tbc1d9
Ptafr
Tmem176b
Gm5763
Tyrobp
E2f2
Acox3
Trim5
Bmyc
Fnip2
Rubcnl
Kctd12
P2rx4
Napsa
P2ry14
Hps3
Slc8b1
Tmem109
Fam26f
Lgals3bp
Npc1
Dclre1c
Dennd1a
Tapbp
Gm8369
Ifi203-ps
Malt1
Bri3
AC132444.1
Apobec1
Tep1
Zfand2a
Lmna
Il6ra
Kif1b
Il21r
Tmem55b
Plcl2
CD82
Hmox1
Lst1
Hgsnat
Pelo
Trim12a
Ifngr2
E2f7
Rgl1
H2-Ab1
Ifi214
Aldh3a2
Rmnd5b
Ncoa1
Ptprj
Atp2b1
Hck
S100a10
Plekhg5
Alox5ap
Hdac4
Usp12
Irf1
Cdk14
Zufsp
Notch2
Synj1
Rps6ka4
Slc26a2
Plekhm1
Tmem50b
Ctsc
Stx7
Gpsm3
Rgs1
Ppp6r1
Sh2b3
Picalm
Iqgap2
Klf13
Map2k1
Gm11189
Clcn7
Tpm4
Pqlc3
Plekho2
Txnip
CD48
Scarb2
Gm7809
Rab3d
Snx20
Parp10
Rnf31
Cyld
Acta2
Hsd17b4
Mknk1
Camk1
Ncor2
Rgs14
Smim3
Mafk
Usp25
Twf2
Arhgef6
Abhd12
Nfatc1
Gnai2
Tiparp
Arl8a
Cndp2
Rnf13
Cnp
Cxcr4
Klf10
Msn
Mpp1
Gcnt1
Haus8
Cyfip1
Pqlc1
Vcpip1
St8sia4
Rhog
Vps37b
Flna
Bet1l
Gnpda1
Gm14608
Fos
Pycard
Plekhf2
Dpy19l1
Stam2
Mvp
Capza2
Necap2
Mif4gd
Gm18588
Tmem189
Add3
Ypel3
Rassf1
Slc6a6
Ifi47
Il17ra
Gm26740
Ehbp1l1
Grk2
Gm12428
Crlf2
Nfkbib
Swap70
Gm23374
Atp6v0d1
Xbp1
Scn4b
Sqstm1
Pstpip1
2610206C17Rik
Pisd-ps1
Ttc7
Fuca2
Gm13077
Skap2
Sdhaf2
Ftl1-ps1
Actr2
Zcchc6
Acaa1a
Emp3
Atp6v1d
Arhgap4
Eif2ak1
Rel
Acot9
Rabep1
Cast
Fes
Abi1
Tacc1
Psme2
Cldnd1
Lamtor1
Ints4
Keap1
Gm5510
Ier2
Ccdc50
Itm2b
Cap1
March5
Rbms1
Gm14000
Bloc1s1
Pias1
Vcl
Commd8
Psme1
Tmod3
Rab8a
Myo9b
Sub1
Snap23
Grb2
Tmem192
Tmem134
Rbfa
Wdr26
Rac1
Smap1
Gm5828
Ndel1
Gngt2
Macf1
Rtn4
Gm15899
Mysm1
Nadk
Il2rg
G6pdx

Ms4a6c
Pfkfb4
Traf5
BC147527
Xaf1
Clec9a
Dhx58
Blvrb
Alcam
Gcnt2
Gpr155
CD86
1600010M07Rik
Slc7a8
Tmcc3
Gm6175
Gm26637
Psap
Gm1966
I830077J02Rik
Hpse
Stat2
Rtp4
Asgr2
4930455G09Rik
Ids
Samhd1
Gm42641
Rgs10
Map3k8
Aldh3b1
Rassf4
Gpr141
Kdm7a
Ak8
H2-T24
Ighj4
Entpd1
Mpeg1
Cd52
Txndc16
CD80
Arl4c
Lrrc25
S1pr2
Irf8
Nostrin
Klrb1f
Pea15a
Bhlhe40
Trim36
Grn
A530040E14Rik
Ighm
Irf5
Tor3a
Stat1
Icosl
Slfn10-ps
Helz2
Nos1ap
Ckb
Samd9l
Irf9
Cytip
Slamf9
Gm6377
Unc93b1
Cd28
Abhd15
Selenop
BE692007
Plekhm3
4833421G17Rik
S100a11
Ciart
Lamp1
Cnbd2
Aoah
Mndal
5430427O19Rik
Ifi30
CD180
Ctsz
E330020D12Rik
Cic
H2-K2
Fam105a
Cyth4
Mef2a
Cav2
Slc2a6
Ddit3
St3gal5
Ifnar2
Gm12854
Tmem176a
Adar
Egr1
Wwp1
Sertad3
Gm9397
Csf2rb
Tor1aip1
Tmsb4x
Fam174a
Prex1
Tnfsf12
Parp9
Klf6
Selplg
Zfyve26
Gm13669
Mkrn1
Sesn2
Tmem150b
Pot1b
Gm15989
C130026I21Rik
Emilin2
Phactr2
Smox
Dusp3
Tpd52-ps
Irf2bp2
Crybg3
B3gnt8
Acer3
Dhrs11
Nfkbia
Plp2
Atp8a1
Gltp
Lpcat4
Relb
H2-DMa
Mkrn1-ps1
Pnpla7
Thap6
Hlx
Camta2
Parp3
Plod1
Stk17b
Dnase1l1
Vhl
Ddi2
Plekhb2
Rnh1
Sft2d2
Rnpep
Gns
Dtx2
Gnaz
Mtf1
Tmem219
Fem1c
Morc3
Stk38
Dna2
Gm4962
Gpr160
Bex6
Coro1b
Pml
Prdx5
Zdhhc7
Trim26
Cdc42se2
Gm8464
Ehd4
Gm21738
Spty2d1
Gm10717
Gm13597
Gm12166
Akt3
Nek7
Arhgap31
Ptprc
Actb
Dpysl2
Crip1
D1Ertd622e
Lamp2
Gm16026
Reep5
Sgpl1
Lipa
Card19
Lrch1
Sidt2
Ppp3ca
Bin3
Gsdmd
Gm8641
Pik3cd
Cyb561a3
Gm17087
Mpc1
Pdxk
Dyrk1a
Elmsan1
Gm12543
Nptn
Gm16372
Kif23
Gm11405
Ep300
Gm12345
Srgap2
Slc31a1
Nagpa
Nfkb1
Foxn2
Gm4750
Rin3
Zfp263
Atp6v0b
Apbb1ip
Cd9-ps
Smim14
Tor1a
Malat1
Atp6v1e1
Spg21
Myo1g
Gm5548
Ankrd44
Uvrag
Atp6v1b2
Zfp36l2
Gm7114
Gm12791
Gm7776
Syk
Vamp3
Cpne3
Gm13453
Mdm2
Cd9
Gosr2
Laptm4a
Ralb
Nub1
Arhgdia
Ccdc50-ps
Gm8399
Mapk14
Sri
Ywhah
Arhgap15
Rn7sk
Runx1
Smdt1
Xiap
Rab1b
Fam49b
Azi2
Azin1
Atp6ap1
Gm2986
Cdc42
Gm12226
Wdr1
Vps51
Erp29
Gm12430
Tspyl1
Far1
Ddx24
Cln6
Chmp1a
Ctsd
Ensa
Gm42786
Mob3a
Gm9392
Rps6ka1
Ptpra
Gm7367
Mvb12a
Taf6l
Oser1
Vapb
Txndc15
Gmip
Szrd1
Mir6236
1700097N02Rik
Arpc2
Smchd1

H2-DMb1	Sema4a	Mthfs	Rap1b
Clcn5	Tpd52	Zfp710	Herpud1
Sema4c	Uba7	Mfsd12	Rassf5
Ero1lb	Map3k3	Gna15	Ptp4a2
Mgl2	Hexb	Exoc3	Psen1
Irgm1	Sgcb	Ccr2	Gm15361
Itgb2	Rarg	Gm10800	Gm8185
Igtp	Rab7b	Rala	Dync1h1
Evi2a	Gvin1	Ythdf3	G6pd2
Rab3il1	Zfhx3	Anxa1	Clta
Tlr6	Ssh1	Gm6695	Snx10
N4bp3	Ppm1m	Pnkp	Nrp1
Fam69a	Mtmr14	Stxbp3	Ccdc12
Gm12892	Gm6169	Arl6ip5	Gak
Tgtp2	Anxa6	Scand1	Gm12164
Ptms	Trp53i13	Samd8	Serp1
Trim30a	Sh3bp2	Trim14	Arpc5l
Pacs2	Ogfr	Taldo1	Gm15534
Tgfbi	Ggh	Tec	

Table 10: Genes downregulated during culture for 5 days in Flt3L differentiation media

432 genes downregulated in main population of 78 *CEBPA* N321D-transduced cells			
Gpc3	Ankrd37	Tmem108	Dtd1
Igll1	6030408B16Rik	Nrxn1	Hist1h3b
Zbtb16	1700012B09Rik	Plod2	Gm12611
Armcx1	Gm11998	Ctsg	Ero1l
Pcdh7	9630013D21Rik	Gm26597	G6pc3
Csgalnact1	Amz1	Tst	Flnb
Lbp	Zc4h2	Clec4e	2810408I11Rik
Ptk7	Vcan	Gpr65	Ydjc
Gfra2	Zbtb34	Gm16104	Mtap
P2rx3	Srxn1	Dab1	Ssbp3
Tspan6	Cdc14b	Col4a5	Tfrc
St6galnac3	Mmp11	Dusp19	Ifi47
8430408G22Rik	Col23a1	Hsd11b1	Rab44
Chl1	Ascl2	CD84	Asph
Mmp13	1700113H08Rik	Clec4d	Gtf2f2
Sema3a	Ube2o	Plscr1	Mmgt1
Ddx4	Gfra1	Spry2	Gm14038
Ms4a3	Nrg1	Depdc1b	Tgfb1
Gpr12	Zxdc	Gm5530	Gm43096
Kit	Ltbp3	Ttll9	Hilpda
Ctla2a	Dab2ip	Gstt2	Chchd10
Pde2a	Cela1	Sirt1	Ap3s1-ps2
Dlg2	Zhx2	Zfp704	Casp7
Hmgcll1	Ighv1-77	Fam46a	Erlin2
Fat4	Polm	Ptger2	Gm16194
Hlf	Pglyrp1	Gm29397	Ybx3
Aebp1	Gimap5	Gm37818	Hist1h2ae
Col4a6	Gm43489	Mpo	Ppid
Rag2	Mettl18	9330175E14Rik	Pfkl
Rbpms2	Hivep3	9330159M07Rik	Rmnd1
Angpt1	Zfp462	Nars2	Chek1
Trem1	Ankrd50	Gfpt1	Gm4705
Ebf1	Klhdc1	Galnt14	Psmd5
Prdm1	Mcpt8	Abcb1a	Gm17149
Ablim1	Il11ra1	Muc13	Endod1
Gata3	Afap1l1	TlCD1	Gnpnat1
Tacstd2	Adgrg7	Fpgt	Vav3
Dntt	Ppic	Adcy9	Por
Ocrl	1700003F12Rik	Zc3hav1l	Ap3s1-ps1
Gab1	Rgs1	Fcgr3	Ap3s1
Sh3d19	Cul7	Pon3	Al314180
Hoxa10	Gm37354	Nfe2	Gm5551
Bmx	Rbpms	Fkbp11	Rasgrp2
Bhlhe41	Tmem132d	Impdh1	Mt1
Slc28a2	Rims2	Prodh	Lsm3
Limch1	Atp11a	Tespa1	Pfas
Gjb3	Hsd17b1	Sapcd2	Suco

Acot1
Car2
Acot3
4930519L02Rik
Pi16
Rsph9
BC026585
Gimap9
Adgrl4
Emid1
Mmp12
F2rl3
Rag1
Efna5
Il6
Itga9
Trem3
Vamp5
Slc22a4
Tha1
Siglecf
Ighv1-23
Igfbp7
Gda
Ctla2b
Csf3r
Zfp27
Gm3470
Cited1
Zfp703
Xrcc5
Fosb
Als2
Trim68
Gcnt4
St3gal6
Igkv4-90
Pomgnt2
Nicn1
Hgf
Six5
Armcx4
4930486L24Rik
B230118H07Rik
Gm38345
Asb1
Optn
Ica1
Tnni3k
Gm12158
Mllt3
Flt1
Slc38a5
Prss57
Plxna3
Zfat
Lekr1
Igkv12-89
Ift81
Elane
Psd3
Ammecr1
Ubxn11
Tet1
Kcnq3
Hist1h3d
Mmp8
AW146154
Galc
Ptprm
Rbakdn
Wdr60
1110019D14Rik
Ubash3a
Olfml2b
Nkg7
Dnajc6
Zfp758
Tmem40
Hist1h1a
Ttc39c
Atp10a
Plcb2
Adgrg1
2310075C17Rik
Hist1h3e
Crim1
Ptgds
Gfi1
Tyw3
Rps6ka3
Gm43431
Zfp934
4933431K23Rik
Plcb4
Gm15581
Gm16897
Homez
Mab21l1
Rgcc
Pygl
Ighmbp2
Arhgap6
Lima1
Ankrd24
Nrp2
Cd34
Clec5a
Rnf125
Slc35b4
Hpgd
Papss2
Epha7
4930555A03Rik
Slc11a1
Parp16
Erg
Car3
Hist1h3g
Meis1
Slco3a1
Dus4l
Mt2
Hist1h2an
Pvr
Dok2
Man1a
Ssfa2
Sipa1l1
Gse1
Smtn
Hist1h2bg
Spred2
Prkar2b
Hist1h3f
Rapsn
Ldlrad3
Gm7805
Ednra
Dtnbp1
Vat1
Ifi214
Kdm1a
Fndc3b
Itga6
Hist1h4n
Pam
Tgfbi
Stap1
Cdkn2c
Abhd17c
Dgat2
Rgs18
Kdelc1
Tmem29
Nrgn
Zswim7
Hist1h1d
Gtse1
Egln3
Gm15658
Hist1h1b
Anxa3
Reps1
Hoxa9
Chsy1
Serpinb1a
Adgra2
Gm9143
Slc17a9
AA474408
Prcc
Hk3
Calcrl
Psat1
Stat4
Hmga2
Steap3
Nipsnap3b
Samsn1
Ccl9
Tsc22d1
Tex30
Sigirr
Cit
Crtap
Arhgap18
Ftsj1
Hist1h3c
Gm16302
Fam60a
Tyms
Nudt1
Hist1h2ai
Atp11b
Gm7389
Rsu1
1810037I17Rik
3000002C10Rik
Aasdhppt
Fam136a
Ppp1cb
Pfdn4
Hmgcr
Mcfd2
Slc19a1
Hist1h2ab
Timm8a1
BC003965
Lmo4
Sptbn1
Ift57
Eya1
Gm13313
Rcn2
Amd1
Gm6104
Ndufa4
Nedd4
Gm14450
Cenph
Mif
Chchd4
Rps29
Cryzl1
Srgn
Gm6065
Gm11349
Orc3
Acyp1
Gm10169
Gm5251
Rpl15
Pnkd
2300009A05Rik
Bphl
Gm14630
Ube2w
Fam162a
Cebpzos
Ebna1bp2
Ergic1
BC028528
Nop10

1942 genes downregulated in subpopulation of 5 *CEBPA* N321D-transduced cells

Mmp13
Tnfrsf1b
Trim16
Gpc3
Ankle1
Flcn
8430408G22Rik
Gfra1
Rps6ka4
Mgat4b
Nfic
Ank
Letmd1
Ankrd13c
Vcl
Man1a
Fam120b
Lrrc45
Rgp1
Myo18a
Cops7b
Htatip2
Mast3
Dhrs4
Traf4
Ddx19b
Kifc1
Phf12
Nploc4
Tmem165
Ptpn1
Sharpin
Gtf2h3
Gk
Wdr75
Dna2

Manea
Plcb3
Urgcp
Osbpl5
Lcmt2
Hlf
Noct
March8
Dus2
Slc25a15
Fam179b
Trim26
Csgalnact1
P2ry2
Hacl1
Ipo13
Ap5z1
Pramef8
Klhl26
Tle6
Faap100
Tmem18
Tango2
Sepsecs
Btrc
3110002H16Rik
Irf1
Spryd4
Acp2
B4gat1
Idua
Ikbke
Mtmr3
Polr3a
Dstyk
Tubd1
Rufy1
Klhl22
Zscan21
Zfp574
Fam160b2
Slc25a44
Ltbp3
Samd10
Usp11
Tk2
Vhl
Gpr160
Igll1
Plekha7
Zgpat
Zscan12
Dennd1c
GuCD1
Scrn2
Det1
Parp10
Mutyh
Fbxl4
Ezh1
Hsd17b7
Zbtb16
Pomk
Ubash3a
Socs5
Pcdh7
Lss
Tbc1d25
Tjp2
Wipi1
Zfp523
Pcp4l1
Tbc1d10b
Mmab
Tsen54
B4galt7
Vps18
Pradc1
Hnrnpll
Scarb2
Dnal4
Rbm33
Hsdl1
Spg11
Tfec
Trappc6b
Pex6
Blcap
Zfp397
Prss57
Socs4
Armt1
Med23
Prdm1
Nf2
Tmem120a
Cited2
Msi2
Ptk2b
Pon3
Ndnl2
Adgra2
Tspyl3
E4f1
Pex3
Myd88
Nbeal2
Ankfy1
Ints3
Tgoln1
4833420G17Rik
Limk1
Acadvl
Chek2
Rab43
Clec5a
Irak3
Mospd1
Rad51d
Rbsn
Kif21b
Zfp608
Aars2
Nfrkb
Osbpl9
Dtx2
Trim11
1810043G02Rik
Klhdc4
Zfp639
Brf1
Fpgt
Fbxw8
Akap8l
Nsl1
Pddc1
Atg16l1
Asb6
Kdm4c
Trem1
Mettl3
Prosc
Erbin
Arhgap19
Srsf4
Plaur
Hsf1
Ifi206
Apba3
Abhd10
Dennd6a
Zfp777
Thap3
Ptpn9
Mbd4
Gpatch11
Slc35f5
Pard6a
Ppm1b
Epc2
Tbl2
Fam53b
Kptn
Mier1
Dnajc6
Clec4d
Cep72
Elp6
Dicer1
Klhl15
Hnrnpul2
Dtwd1
Stim2
Fam216a
Dclre1c
BC029722
Dcp1a
Tnrc18
Ercc6
Cntln
Kdelc2
Optn
Isoc2b
Antxr2
Sbno2
Inf2
Tnip2
Tnip1
Mlst8
Tmub1
Tsen15
Gmeb2
2410004B18Rik
Emilin2
Snapc5
Dab2ip
Tbc1d7
Arhgap32
Myo9b
Ift74
Tbc1d1
Rtf1
Rsph9
Ythdc1
Paox
Cntrob
Slc35a3
Xpa
Kpna6
Mia3
Wasl
Coa4
Kdm5c
Anxa9
Hck
Kat6a
Tbp
Mcm3ap
Snapc1
Zwilch
Gm8546
Cdk19
Cercam
Clec4e
Cdkn2aip
Gm9774
Gm13803
Rasgrp2
3000002C10Rik
Zc3h7b
Mmp8
Gm37784
Pck2
Itga5
C2cd5
Snord55
Frrs1
Gabpb2
Pdhx
PdCD4
2310009A05Rik
Gm12791
Hgf
Nup50
Smtn
Jarid2
Slc35b4
Fbxw11
Paf1
Phkg2
Hspa4
Gm2810
Dirc2
Supt6
Rptor
Helz
Stim1
Uap1l1
Gm1969
Gm29462
Rapgef1
Ppcdc
Aurka
Rrp8
Zhx1
Rrp7a
Zcchc6
Wdr5
Bhlhe41
Gm7162
Bcl7b
Gm15530
Vprbp
Mcph1
Gm8818
Yif1b
Pik3ip1
Rbms1
Gm13517
Mt2
Ska1
Nr2c2
Ing1
Scamp2
Ate1
Mapkap1
Rabepk
Scaf1
Ccdc43
Akt1
Fasn
Pprc1

Sdccag8
Wrap53
Spice1
Dhx34
Ddx59
Ilvbl
Nrp2
Slc25a25
Gfra2
Zswim1
Daglb
Mon1a
Bbs2
Slc2a3
Brat1
Mettl8
Insig2
Pdik1l
Hps6
Mr1
Gsg2
Ppp1r16a
Cyp4f13
Cep131
March2
Glipr2
Gclc
Fam174b
Armcx1
Tmem110
Fam185a
Alg12
Zfp729b
Spryd3
9130011E15Rik
Pgpep1
Slc38a7
Slc37a1
Usp28
Pomgnt2
Nedd9
Pex12
Aifm2
P2rx3
Trp53i13
Cyb5rl
Lima1
Snrk
Sgk1
Snupn
Hook2
Cdk5rap1
Cep89
Alkbh4
Osbpl7
Hcfc2
Fut4
Rsad1
Aldh7a1
Pim3
Slc2a8
Unc45a
Ccdc77
Myadm
Fam160a2
Lbp
S1pr3
Tmem268
Baiap2
Hyal2
Sh3bp5l
E130307A14Rik
Pars2
Map4k5
Cast
Nfatc3
Oip5
Fem1a
DHRSX
Tyk2
Cdk5
Tmem214
Hibch
Dhx29
Gpr180
Fam3a
Nol6
Ctso
Man1c1
Trmt5
AC149090.1
Zfp869
Etv5
Rint1
Mllt10
Dgcr14
Sgms1
Herc2
Mab21l1
Phf14
Azi2
Brpf1
Eme1
Pigc
Gas2l3
Lrp6
Utrn
Nr2c1
Dennd4b
Caml
Rtn4ip1
Ppfia4
Gyg
Ints6
Bloc1s5
Tceanc2
Secisbp2l
Serpinf1
Pop1
Golim4
PdCD11
Morc2a
Exosc4
Trappc8
Qpctl
Cenpo
Nphp1
Rnf123
Bicd2
Zfp692
Phf7
Dguok
Zzef1
Mgat4a
Sft2d2
Hps5
Irx2
Tti1
Trappc3
Commd5
Clasrp
Chchd6
Fam195a
Klf2
Lztr1
Stat4
Champ1
Tcta
Atat1
Cdc37l1
Zmym4
Prkag2
Pip5k1a
Rnf125
Lrif1
H2-Ab1
Kif18b
Arnt
Pura
Arrdc3
Kdm2b
Aco1
Nars2
Tle3
Insig1
Troap
Oxnad1
Trmt61b
Hps4
Pigl
Gtf2ird2
Polr3gl
Rnf213
Prune2
Fam219b
Map1lc3a
Xrcc2
Zfp668
Nfkb2
Ubxn2b
Papss1
Ankra2
Cdk13
Lin37
Scyl2
Smg8
Relt
0610009B22Rik
Grb2
Plekhg3
Arid5a
Spag5
Amz2
Nt5dc1
Oxld1
Fkrp
Tsr1
Mtf1
Rbm45
Erich1
Rhebl1
Tmtc3
1700109H08Rik
Pqlc1
Dhx38
Ndst1
Ipo4
Irak1bp1
Kmt5c
Ahi1
Fdxacb1
Dhdh
Csgalnact2
Slx1b
Twsg1
Gm19680
Dlg1
Prepl
Zfp369
Cenpe
Ifitm1
Actn4
Mapk14
Tpm1
Acot7
Mrps5
Dnajb12
Yif1a
Golga5
Pdpk1
Kdm6a
Dctd
Naa60
Tagap1
C330027C09Rik
Vav3
Gm6158
Cdk16
Fam126b
Arhgap31
Gm6623
Ddx49
Ncf4
Asl
Gm15800
Sppl2a
Raly
Sdc1
Cnot1
Dhx36
Mthfd1l
Nudcd3
Lonp1
Dctn2
Micu2
Arid1a
Nomo1
Suco
Stx8
Cul2
Cdca7l
Sptbn1
Mogs
Fbl
Map2k3
2610206C17Rik
Ube2g2
Numa1
Prkd3
Ptcd3
Ebp
Setd2
Guca1a
Hist1h1c
Nr1h2
Mdm2
Tap1
Fam192a
Celf1
Kit
Haus4
Tomm40l
Ltbr
Cluh
Hnrnph2
Wdr82
Arfgap2
Usp22
4921524J17Rik
Gm15519
Os9
Mettl1

4933404O12Rik
B3galt6
Wbscr27
Wdr91
Txnrd3
Cog3
Coq8b
Ndor1
Nuak2
Appl2
Gem
Gba2
Slc25a29
Tmem237
Vps8
Taf8
Terf2ip
Slc35e1
Arrdc2
Gbp7
Zfp775
Tmem185a
Recql4
Git1
Irf2bp1
Tepsin
Cbx2
Aff1
Zfp407
Plscr4
Hace1
Dnajc14
Recql5
Syne2
Tmtc4
Arhgap6
Tbc1d10a
Mmp12
Rab44
Dntt
Plod2
Angpt1
Emid1
Ssfa2
Sema3a
Slc11a1
Chl1
Gaa
Sapcd2
Muc13
Por
Ppm1f
Mllt3
Idh1
Ddx4
Galnt14
Pde2a
Ccnf
Emilin1
Gys1
Stxbp2
Pfkp
Rgs18
Papss2
Preb
Bin2
Lpcat2
Stk11
Reep4
Zfp703
Aebp1
Erlin2
Flnb
Atp5s
Arfgap3
Ash1l
Mtmr12
Gzf1
Mul1
Tacstd2
Mto1
Hgh1
Pdlim7
Rb1cc1
Zfp157
Polr3d
Poc1b
Nagk
Ppp2r5e
Pcnx
Selenoi
Poli
Gm20342
Rffl
Fgd4
Rfx7
Nsun4
Hdac5
Unc13d
Rpusd4
Rnpepl1
Entpd6
Ipo8
Gpatch1
Rnf145
Panx1
Cep57l1
Cdc25c
Lactb
Supt7l
Kctd10
Rbm28
Rraga
Plod1
Akap9
Yipf2
Map3k4
Iws1
Mcm9
Stat5a
Tmc6
Pik3ap1
Gramd3
Pus7l
Epb41l2
Traf3ip3
Coq7
Aldh4a1
Nrp1
Faap20
Pi15
Itgb3
Dnajc13
Rap2c
Gca
Nfat5
Golph3l
Dvl2
Gpn2
Cpped1
Wrnip1
Rgs1
Sumf2
Ikbip
Sphk2
Mccc2
Primpol
Sec24b
Tbc1d22b
Ube2d-ps
Vezt
Mospd2
Dus4l
Gm13328
Cluap1
Arl10
Tpcn1
Sos2
Scrib
Usp36
Cacfd1
Sik2
Abl2
Dse
Hps3
N4bp2l1
Rag1
Csf2rb
Fam118b
Wwc2
Dcaf4
Rab5b
1110038F14Rik
Dhrs11
Gm12454
Gfpt1
Pspc1
B4galt3
Prelid2
Gadd45a
Cbfa2t3
Mesdc1
Fez2
Mterf3
Rbm6-ps1
Ptpn4
Ccdc88c
Rbbp6
Nfkbil1
Focad
Tmem156
Sfxn2
Psd3
Dennd1b
Mgme1
Shprh
Upf1
Fam76a
Tex264
Tnks
Mapk3
Gin1
Dusp19
Prr11
Nubp1
Fra10ac1
Camk2d
Gcc2
Cyb561a3
Afg3l2
Pcyt1a
Endog
Thap4
Vps13b
Xndc1
Osbpl2
Cenpl
Tal1
Zfp825
Hk1
4930453N24Rik
Nrgn
Gm26917
Fads1
Nkiras2
Sigirr
Cpsf7
Setd5
Notch2
Cdc34
Acot2
Nosip
Dapp1
Ankrd11
Gtf2i
Tbl3
Cdc20
Dusp6
Epha7
Dnase2a
Zmpste24
Pik3r1
Mrpl58
Fam172a
Crtap
Dok3
Gm10051
Itgal
Cdk11b
Ap3s1-ps2
Pxn
Cdc27
Rrp9
2410002F23Rik
Acin1
Gpbp1l1
Fam46a
Eif2d
Rcn2
Plscr1
Csnk2a1
Anxa1
Adprh
Tctex1d2
Bckdha
Cnot9
Mmadhc
Ppp1r8
Ubr4
Jagn1
Gm8010
Gm5530
Yeats4
Suv39h1
Lyz2
Srprb
Btk
Itpr1
Acap1
Selenom
Tmed7
Gm38078
Prr14
Tmem39a
Impdh1
Gm7389
Anxa3
Dlst
Pdia4
Nudt3
Apobec3
Gm5499

Lmo4
Ikbkap
Trabd
Zfr
Gpr12
Smg5
Tmed1
Nat10
Ankrd27
Prkar2b
Nfe2
Acot1
Anapc7
Vps26b
Armc5
Mfge8
Wars
Ctla2a
Ednra
Skiv2l
Polrmt
Miip
Sart3
Dhodh
Gstz1
Atp13a1
Hmgcll1
Pcyox1
Col4a6
Gorasp1
Ptprm
Hilpda
Chid1
Dlg2
4930427A07Rik
Man2c1
Cant1
Orai3
Wwp2
Pitpnm1
Stat6
Clip1
Narf
Fut7
Pml
Inpp5b
Fat4
Tgfbi
Dmtf1
Slc9a9
Ttll12
Ggta1
Aldh3a2
Eif2b2
Fastk
Sipa1l1
Hoxa10
Gch1
Slc39a11
Rdh11
Safb
Col4a5
Clpx
Fbxo18
Ciao1
Slc17a9
Ubtd1
Ldlrap1
Vps39
Cyth1
Jmjd6
Bag5
Bet1
Mctp1
Pcsk4
Ocrl
Ppp1r14c
Pex16
Ttk
Pex1
Hmbox1
Hmg20b
Ncoa6
Tmem218
Cipc
Ankrd28
Cox18
Zpr1
Cpne2
A230050P20Rik
Abhd6
Pecr
Manbal
Yeats2
Tpgs1
Bsn
Nprl2
Hmces
Kmt5b
Med26
Vcpip1
Zfp282
Pstk
Gab1
Wdyhv1
Ccdc18
Nbas
Sugp2
Mrpl45
Fam188a
Usp42
St7
Aldh6a1
Grpel2
Sh3glb2
Ccdc191
Tjap1
Mical1
Uros
Tirap
Sh3d19
Snap29
Map2k7
Parp16
Csnk1g3
Bcor
Polr3c
Birc2
Klhl7
Thap7
Ptch1
Agps
Tmem199
AU040320
Zbtb41
Phyh
Arap1
Tmco6
Nfe2l1
Sbf1
Zfp646
Trem3
Arsa
Gnaq
Serinc5
Prr3
Dusp12
Vamp5
Praf2
Gab2
Gigyf1
Tmem259
Sun1
Thg1l
Gm18284
Tfb1m
Phf8
Aida
Dedd2
Bud13
Ccdc186
Fam49a
Slc22a4
Purb
5730480H06Rik
Sirpa
Bahcc1
Furin
Ermard
Lamc1
Gm6493
Brwd1
Wdr62
Sfmbt1
Igsf6
Hist1h1d
Arhgap10
Srf
Plcg1
Ighv1-23
Cdk5rap2
Inpp5f
Scly
Nudt13
Actr6
Trak2
Pafah1b3
Gse1
Ick
Ints8
Sufu
Csnk2a2
Gapvd1
Rpia
Kcnj2
Ncor2
Rnf216
Gm5874
Slc25a12
Ppp1r9b
Adgrl4
Ubqln1
Slc35e2
Gm12166
Map2k4
Fbxw2
Strn
Fam26f
Ap1g2
Ranbp10
Plcb4
Zfp81
Gm7512
Dpy19l1
Zbtb45
Depdc1b
Dctn1
Gm9517
Yod1
Mia2
Rbm5
Ahsa2
Pdxdc1
Zdhhc20
Cd72
Stom
Usp9x
Smc2
Ccar1
Ift20
Atp6ap1
Lpcat3
Myc
Leng8
Hmgcs1
Ifrd2
Mars
Tpp2
Ube2w
Dpp3
Pgd
Tm9sf4
Gm2a
Pigu
Elk3
Alkbh5
Herpud1
Timm17a
Tln1
Mir5099
Tacc3
Acly
Nipbl
Ewsr1
Arhgap18
Pwp1
Rnf181
Gnpat
Dcaf7
Xpc
Tmem183a
Gm7556
Stk4
Gm7676
Map1lc3b
Vasp
Dld
Tmx2
Cd34
Tmem242
Tbcb
Gm9923
2810403A07Rik
Eprs
Zfp622
Glud1
Fbxo9
Eif3d
Arl6ip4
Cdca8
Dcun1d5
Huwe1
Gm12716
Cd33
Nedd4
Ddx46
Actr1a
Bcat2
R3hdm1
Ncstn
Rpl28
Arf5

Glt8d1
Sipa1
Dhx30
Mettl2
B230219D22Rik
Neu1
Xbp1
Tbc1d17
Akr7a5
Rprd1b
Fam63a
Tst
Ebf1
2510039O18Rik
Gla
Opa3
Dgcr2
Prodh
Prpf38b
Atad3a
Akr1b10
Keap1
Atf7ip
Ints5
Eml3
Ccdc71
Kif2c
Rbpms2
Gne
Btbd10
Poldip3
Letm1
Mrps31
Meis1
Rere
Rad9a
Myo1c
Cpt1a
Telo2
AI467606
Hist1h2bc
Pde4a
Rwdd4a
Pvr
Kansl3
Nudt9
Ampd2
Tmem40
Ccnl2
Leo1
Agpat4
Afap1
Mettl16
Gm17018
Gcdh
Wdr73
Gpr65
Naglu
Gga2
Mavs
Usp21
Hmg20a
Cpt2
Eaf1
Mvd
Ablim1
Rbpms
Gjb3
Kctd5
Mcat
Mars2
Wdr3
Arhgap11a
Rmnd5a
Gata3
Xxylt1
Gna15
Ptger2
Map2k6
Fbxo28
Ralbp1
Nusap1
Atg2b
Mfng
Lgals3bp
Mdm1
Snip1
Pnpo
Amz1
Fam96b
Tmem68
Plpp2
Slc9a1
Dmap1
Ccdc61
Ctdsp1
Pelp1
Fam98a
Ing2
Ifi204
Tada2a
Mettl5
Dffb
Msrb2
Prpf3
Erg
Asb3
Arhgef3
Cwc22
Extl3
Tiam1
Rassf1
Hinfp
Gga3
Polr3b
Rnf214
Klhl12
Ipp
Rnf166
Naa25
Dhcr24
Mettl7a1
Fadd
Ggact
Stau2
Tgfbr1
Rpap2
Tespa1
Xrcc1
Bet1l
Cutc
Fnbp1
Rdh13
Mfsd12
Fbf1
Pus10
Cep120
Mpi
Nemp1
Slain2
Tspan5
Slc35a2
Acox1
Ubp1
R3hcc1l
4930523C07Rik
Galc
Ptafr
Gm20696
Fcgr3
Arl8a
Plk-ps1
Chd3
Pcnx3
Htt
Nkg7
Polr1a
N6amt1
Cyb5r1
Gm21738
Safb2
Necap2
Plxnd1
Glul
Pcdhgc3
Fam126a
Stxbp5
Lsm10
Tarsl2
Rabl3
Akirin2
2310075C17Rik
Gm3470
Smad5
Ring1
Tob2
Gtpbp1
Cox11
Hist1h1a
Fchsd2
Kif18a
Gemin8
Eps15
Pced1a
Zfp277
Dhx33
Naa40
Tm7sf2
Prkaca
Ugt1a7c
Rexo4
Kat5
Gpr89
Actr8
Dolpp1
Lrrc24
Rapgef2
H2-Q5
Dus3l
Rnf44
Asph
Hdgfrp2
Tmem184b
Igkv4-90
Fam222b
Otud5
Otulin
Sptan1
9330159M07Rik
Mid1ip1
Tmem203
Gm5578
Midn
Zfp36l1
Jmjd8
Hipk2
Sec24a
Crtc3
Adrm1
Rp9
Pdcd6ip
Tfrc
Gm8806
Rpl7a-ps3
Raver1
Fip1l1
Paip2b
Mrps12
Pnn
Trf
Nisch
Ppig
Ep400
Fubp1
Ctsg
Dtnbp1
4930519L02Rik
Gm6560
Ubald2
Hoxa9
Cd63
Aldoart2
Eno1b
Thrap3
Man2b1
Hk3
H2-K1
Pdap1
Pold2
Ctsd
Brd3
Prrc2c
Gm10819
Hdlbp
Calr
Gm5814
Srrm2
Rrbp1
Gm9826
Ccdc124
Srp72
Rnh1
Sf3b2
Ing4
Eya1
Nipsnap3b
Mff
Adgrg1
Pkm
Iqgap1
Sucla2
Smarca4
Ero1l
Gm42418
Prdx5
Gm5276
Nfu1
Dctn3
Lyl1
Sdf2l1
Dnmt1
Gm5424
Hist1h1e
Eno1
Mrfap1
Ddx1
Gm13612
Slc39a7
Cd63-ps
Sec61a1
Ctsc
Hspa5

Hdac10
Srbd1
Fam175b
Cep250
Lig3
Dhx16
Shkbp1
Ube2c
Sec22a
Chd1l
Wls
Tmem161a
Ccdc137
Tbc1d14
Ngrn
Txnrd2
Cry1
Card9
Atr
Sel1l
Msmo1
Sec23a
Slc25a24
Triobp
Hyou1
Gbe1
Btd
9630013D21Rik
Dnase1l1
Cd200r1
2610301B20Rik
Gtf2h2
Fance
Pgghg
Bscl2
Spast
Cog4
Mvk
Agpat1
Cstf1
Tvp23b
Ccdc86
Wdr24
Jak3
Tmem184c
Ubl7
Creb1
Plekho2
Lgals8
Dhx35
Atp6v0a2
Arhgap27
Zfp180
Mipep
Slc35a1
Cd27
Wdr6
Akt1s1
Cog1
Nrf1
Banp
Wrap73
Ppox
Fem1b
Dnaja3
Taok1
Diaph1
Srd5a3
Scamp3
Kcnab2
Pggt1b
Dnmt3b
Rexo1
Irf3
Ttc17
Add3
Rnf31
Bmx
Mbd1
Rbm6
AI597479
Slc19a2
Car9
Stxbp4
Tm2d1
Qsox1
Bcr
Ubash3b
Mov10
Adprm
Trim8
2700097O09Rik
Mtmr14
Vgll4
Tmem19
Tnpo2
Rhoq
Gm9144
Ckap4
Ppp1r10
Eral1
Stxbp3
Kdm5b
Ctdsp2
Zyg11b
Dhx40
Smcr8
Orc1
Traf2
Tsc1
Khnyn
Ncoa2
Phrf1
Rnf41
Gpatch2l
Aspm
Rwdd2b
Ankrd26
Pemt
0610009O20Rik
Vmp1
Wash1
Gatb
Tmco4
Lclat1
Ppp4r1
Ccnb2
Pla2g6
Mrs2
Rabggta
Nfyc
Cstf2t
E2f6
Slc28a2
Ddx54
Mknk2
Ipmk
Mettl13
Snx15
Usb1
Zmym5
Mvp
Ggh
Ppp1r12c
Iqce
Fuz
Fam3c
Ywhag
Sugp1
Zdhhc21
Prrc2b
Mettl4
Gm20604
Abi3
Tbk1
Mtfr2
Pan2
SmarCD1
Itpk1
Prcc
Sla
Stk11ip
Gm12158
Csf2ra
Bpnt1
Npepps
Dmxl1
Kif4
Trim14
Pam
Ano6
Zfp800
Plk1
Gnpda2
Spry1
Casc3
Gm5303
Msl1
Setd7
Tcof1
Mrps30
Abhd2
Krt10
Eif4g1
Gm28417
Foxm1
Rab28
Slc19a1
Sft2d1
Ube2h
Rin3
Tssc1
Gm18588
Tmbim1
Pcgf6
Sars2
Tmem5
Aldh16a1
Lrpap1
Ppt2
Gm4258
Ripk1
Mbip
Gm13880
Gm10801
Ncoa3
Rasa3
Igkv12-89
Rangap1
H2-M3
Elmsan1
Rcbtb2
Ndrg3
Faf2
Gga1
Fntb
Arhgef12
Ppp2r5a
Zmynd11
Rbmxl1
Foxp1
Stoml2
Ap3s1-ps1
Tram1
Uba1
Gm4735
Plscr3
Ndufs2
Dhx9
Gm9575
Ccl9
Gm4784
Ube2i
Gm11964
Eif4a2
Edf1
Rab11b
Myb
Gm7180
Anp32b-ps1
Ccnd3-ps
Phb
Rab11b-ps2
Arl1
BC031181
Drg1
1110008F13Rik
Lsm6
Prtn3
Rnaseh2c
Calr-ps
Rbm25
Malat1
Lars2
Hmgb3
Sdhc
Elavl1
Vim
Hsp90b1
Lman2
Pfdn6
Chchd10
Pebp1
Tmem59
Gm4737
Psmc1
Ap2m1
Gm12611
Psmd7
Cuedc2
Surf4
Gm13509
Aldoa
Anp32b
CD164
Csde1
Npm3-ps1
Anp32e
Gm6517
Pdia6
CD47
Psat1
Akr1b3
Cox7c
Tmbim6
Eif4h
Gm9824
Aldoart1
Selenok
Nono
Pgk1
Phgdh

Fam207a
Dcaf17
Arntl
Rag2
Ice2
Pum1
Agtrap
Smyd3
Klc4
Wdr83
Strip1
CD84
Armc8
Decr1
Ado
Rpap1
Gm16194
Slc25a19
Tmx4
Tusc3
Ngly1
Chek1
Wbscr16
Zfp266
Dis3l
Tbc1d19
Rgs14
Selenbp1
Pgm3
Aim2
Sec14l1
Ska3
Urb2
Vipas39
Ddx11
Pqbp1
Katnb1
Tusc2
Golga4
Cby1
Rnpc3
Lpcat1
Ip6k1
Nudt22
Kctd1
2610020H08Rik
Nus1
Scmh1
Zfp146
Dpy19l3
Rapgef6
Cebpg
Rnmtl1
Lman2l
Kif14
Trdmt1
Il3ra
Gpank1
Pgam5
Ttc39c
Exo5
Dbt
Tapbpl
S1pr4
Vars2
Ddx31
Trp53bp1
Cmas
Fbxo6
Ighmbp2
Vegfb
Tns3
Vkorc1l1
Ccdc94
Kmt2a
Kctd13
Mapk8ip3
Ddx55
1810044D09Rik
Pmm2
Supt20
Cxxc5
Ppwd1
Gm23130
Sfswap
Scap
6330562C20Rik
Saal1
1700003F12Rik
Sumf1
Fzr1
Spsb3
Slc25a1
Pisd-ps1
Ino80dos
Ccdc69
Mtmr6
Klf6
Heatr6
Bod1l
Gm12397
Ehbp1l1
Hint2
Cmtr1
Oxr1
Wdfy2
Gm9200
Snord100
Pycr2
Mark2
Gm7856
Arid5b
Net1
Gm7334
Gm3436
Myh9
Cdkn2c
Cct2
Rsu1
Manf
Pgam1-ps2
Tkt
Itm2b
Gm7901
Rpn2
Etfb
Arpc1b
Rnaset2b
Ubc
Rbm8a
H2-D1
Serpinb1a
Prdx6b
Rbm39
Gm5637
Glrx3
Gm12669
Gm12020
Atp5a1
Pgam1
Ucp2
Cct6a
Tomm20
Gm6368
Srgn
Hsp90ab1
Clic1
Park7
Capzb
Ldha
Tdpx-ps1
Selenoh
Serbp1
Vdac2

1340 genes downregulated in EV-transduced cells

Gjb3
Cd99l2
Slc25a29
Thsd1
Ptgir
Ppic
Zfp775
Gm807
Gpc3
Adgrl4
Slc45a3
Gfra2
Igll1
Dntt
Gfra1
Angpt1
Emid1
Cand2
Chl1
Slc43a1
Cul7
Hmgcll1
Ptk7
Hgf
Gpr12
4930519L02Rik
Ptprm
Ablim1
Rbpms2
Col4a6
Pdss2
Gm12158
Gm21807
8430408G22Rik
Fndc3b
Gm4349
Cd302
Jpx
Dab1
Adat1
Gm37666
Zfp566
Mt2
Galc
Vpreb1
Vangl2
Mrgpre
Dennd5b
1700003F12Rik
Zdhhc24
Gdf11
Slc18a1
Kcnq5
Pla2g6
Tmem40
Il6
Zfp507
Prkca
Tnfsf13os
Prune2
Chchd4
Pla2g4a
Ift57
Fer
R3hcc1
Gm24452
Zfp101
H2-Q7
Zfp512b
Gm5385
Snord22
Itga6
Snx27
Tspan13
Gm5124
Aldh18a1
Mfsd13a
Bzw2
Phgdh
Zak
Acaca
Unkl
Srgn
Bambi
Ttc37
3110082I17Rik
Zfand4
Mif
Fancd2
Gm16379
Fubp1
Ampd2
Utp11l
Gfm1
Ptprcap
Gm11847
Gm14109
Nme1
Ndc1
Gm5069
Mrps28
Znhit3
Trim28
Ifrd2
Ripk3
Cdk7
Ndufaf2
Cad
Endod1
Chchd5
Gm6640
Nme6
Wdr77
Pus10
Gm11914
Gm7781
Trappc13
Pgam1-ps2
Rps2-ps10
Gm12403

Rag2
Aebp1
Dlg2
Hoxa10
Ica1
Fat4
Muc13
Sh3d19
Ebf1
Pear1
Col4a5
Gimap6
Rcn1
Ddx4
Gata3
Csgalnact1
Sema3a
Car2
Hlf
Galnt14
Tacstd2
Mcpt8
Limch1
Aldoc
Plod2
Erg
Ms4a3
Ctla2a
Psd3
Nefh
Ubash3a
Cd27
Ppp1r14c
Vamp5
Tnfrsf26
Dnajc6
Car3
Tmem108
Gab1
Arhgap6
Plscr4
Pcdh7
Slc28a2
Armcx1
St6galnac3
Atp11a
Ptgds
Rag1
Ltbp3
F2rl3
Afap1
Fam160a2
Pecr
Dync2li1
Cbx2
Tnik
Rhoj
Wbscr27
Ednra
Acot3
Rgcc
Mab21l1
Gm15261
C030034L19Rik
Itga9
Oxnad1
Armcx2
B230118H07Rik
Kit
Zfp239
Lbp
Gstt2
BiCD1
Xpnpep3
Qtrtd1
Zfp934
G2e3
Dnmt3b
Brinp3
D130058E05Rik
Clstn1
Fam35a
Ppfibp1
Ttc39c
Prss57
Tst
4732463B04Rik
Simc1
Gm8113
Fancb
Unc13a
Mmp13
Ppp1r13b
Gm9847
Uevld
Camsap1
Ube2cbp
Katnal1
Txnrd3
Col23a1
Gm30074
9930104L06Rik
Taf1a
Cep41
Mnd1-ps
Pomk
Klhl12
Pde8a
2610008E11Rik
C230034O21Rik
Msantd3
Nav2
Selenon
Mnd1
Spry2
Kcnq3
Tspan32
Zfp40
Hyal2
Zfp748
Pdpr
Kctd4
Trem3
Fam114a1
Dach1
Gm16104
Pus7l
Ankrd13b
Sapcd2
Hmgb3
Mlh3
Pced1b
Layn
Kremen1
Ctnnal1
Thy1
2610306M01Rik
Ahdc1
Zhx2
Gca
Smim24
Chchd10
Zfp932
Cep97
Stxbp5
Yae1d1
NudCD1
Guf1
Tma16
Gm13998
Mcph1
Spint2
Zbtb38
Kctd1
Orc2
Gm10169
Gpatch4
Zfp672
Wdr6
Bphl
Gm4353
Gm5455
Ppp1r14b
Nudt7
Gm14328
Grhpr
Gm5946
SCD1
Stat4
Rpp14
Qsox1
Ints3
E130308A19Rik
Strbp
Kti12
Gm43297
Dnajc12
Suv39h2
Ddx1
Samsn1
Mgat5
Wdr55
Cd24a
Gclm
Bmi1
3000002C10Rik
Gm7332
Gm1818
BC017158
Mns1
Plppr3
Ftsj1
Fam92a
Tspan2
Ints6
Plpp5
Pus7
Galk1
Akap1
Srm
Spout1
Sergef
Gemin4
BC055324
Gm10020
Nipsnap1
Selenbp1
Gm43096
Yod1
Abcb8
1110008L16Rik
Eya1
Adgrg3
Itgb3bp
Gm6652
Bcl7a
Gtf2f2
Kansl1l
Tespa1
Gar1
Oaz2-ps
Gcsh
Hadh
Mrpl50
Gm7666
Pgk1
Ybx3
Gm16074
Ruvbl1
Aasdhppt
Snhg4
Gm9143
Abhd11
Rps6
Slc7a6
Gm16433
Selenom
Bcap29
Ubxn8
Comtd1
Pi4k2b
Sssca1
D16Ertd472e
Mrpl38
Gm13700
Pom121
Chordc1
Nudt5
Ntmt1
Ergic1
Taf5l
Eef1g
Mrps31
Por
Blm
Pum3
Anks1
Gm8522
Gm15484
Gm6501
Rrs1
Atg12
Rpl36
Ubfd1
Gm26917
Smyd2
Top3b
Fbxo9
Gm11896
Gstp-ps
Paip2
Gm4735
Clpp
Mcm2
Fkbp3
Stam
Mvb12b
Gm17837
Gm4366
Ddx56
Stk26
Cluh
D5Ertd605e
Cenph
Pabpn1
Mir5099
H2-Q6
Gaa
Eif1ax
Lman1
Ipo11
Idh2

Bmx
Adgrg1
Gm10384
Traf4
Fgf10
Ocrl
Map9
Nrgn
Pglyrp1
Mc5r
Hmga2
Slc22a4
Bhlhe41
Dpy19l3
Lima1
Gimap9
Tal1
Rpusd2
Bcat1
Gm3470
Hoxa9
Plcb4
Rai14
Dab2ip
Phlda3
Epha7
Abhd6
Igkv4-90
Col19a1
Cbx6
Ptger2
4930555A03Rik
Efna5
Spry4
Gimap1
Ighv1-23
Ccdc15
Plxna3
Mmp11
Pi15
Mettl15
1700030K09Rik
Mzb1
Gcnt4
Igfbp7
Dnajc27
Ahi1
Lsmem1
Aasdh
2810002D19Rik
Gsta4
Htra3
Spns2
Tspan6
Zfp783
Gata2
Mpo
Kcnj2
Maged2
Mei4
Jmjd4
Kcnip3
Senp8
Amer1
9630013D21Rik
Gda
Stxbp4
Tnk2
Socs5
Sort1
Gm26580
Cngb1
Xk
Gm6104
Il31ra
Prelid2
Larp1b
6330562C20Rik
Ube2o
1190007I07Rik
Exo5
Hist1h1a
Spice1
Hyls1
Cmss1
Mtx3
Tnnt1
Akap7
Slc35d3
Osbpl7
Etfbkmt
Ccdc181
Gm9256
Gas2
Bhlhb9
Sirt5
Tfrc
Slc37a1
Gm37818
Capn3
Gm16341
Gm37422
Prkar2b
Gemin5
Impdh1
B230369F24Rik
Gm37510
Rpusd3
Gfi1b
Aak1
Dok2
Abhd17c
Ippk
Gtf3c4
Lins1
Angptl4
Urb2
Rrp12
Dctd
Zfp219
Mettl4
Sh3rf1
Rbak
Zfp979
Gm37124
Dtx4
Exoc8
Slc35b4
Gse1
Utp14b
Tceanc
Etv4
Tmem67
Micall2
Wdr5b
Lrrc20
L2hgdh
Gm29397
Bag2
Kif17
Mroh2a
Crnde
Sox4
Poli
2610301B20Rik
Focad
Pbx4
Gm5847
Gm5864
Smpd4
AI987944
Pam16
Snord55
Gm23246
Polr1b
Dirc2
Ccdc167
Gm15659
Pitrm1
Fam136a
Rrp1b
1110004E09Rik
Gm17494
Grwd1
Fam216a
Lyar
Gm14769
Sord
Rint1
Kcnq1ot1
Enkd1
Gm14450
Hist1h2bj
Nup37
Gm13502
Reps1
Fam162a
Trp53bp1
Pdlim1
Cdkn2aipnl
Gm17066
Mettl1
Abcb6
Gm5703
Nolc1
Snhg15
Spata24
G6pc3
Acot2
Utp20
Gm4737
Cdk6
B3galnt2
Las1l
Wdr12
Cerkl
Dnph1
Lmo4
Jmjd8
Rai1
Nt5c3b
Nob1
Gm13337
Rasgrp2
Exosc2
Trub2
Flnb
2900026A02Rik
Nme4
Ahcy
Wdr75
Eef1e1
Dnaja3
Mdn1
Cdkn2c
Gnpnat1
Sptbn1
Xpo7
Nup133
Psme3
Mthfd1
Mesdc2
Borcs7
Gm13719
Emilin1
Gstp1
Gm7556
Mycbp
Gm17709
Ninj1
Cdk4
Rfc4
Cbx5
Hmgcr
Gm8173
Tia1
Srek1ip1
Ftsj3
Icam2
Gm10819
Eno1
Asf1a
Dhx30
Cit
Pebp1
Gm13368
Gm9824
Bola1
Gm12115
Trap1
Hddc2
Rpl27
Nomo1
Prep
Rpl31
Tpi-rs11
4921524J17Rik
Rps12
Klhdc2
Pold2
Gm6447
AI314180
Tmem254c
Tomm5
Gm5593
Ktn1
Gnl3
Aimp2
Gm5611
Gart
Gm13552
Tfap4
Nsun2
Nop56
Tpi1
Haus1
Clec4e
Oaz2
Tufm
Mrto4
mt-Nd2
Ptpn7
Drap1
Nifk
Mrpl15
Tyms
Gm37201
Gadd45gip1
Rnf220
Tmem70
Gm7244
Trnt1

Arpp21
Gm38391
Dennd6b
Rtn4r
Map7d2
Olfml2b
Pde2a
Cited1
D630045J12Rik
4930509E16Rik
Nck2
Tha1
2310075C17Rik
Mmp14
Tarbp1
Abcd2
Ctxn1
Camkk1
Slc25a43
Rapgef3
C920006O11Rik
Gm20696
Rnf125
1700012B09Rik
Mgmt
Myo19
Fam220a
Gem
Taco1
Car8
Armcx4
Slc38a5
Zfp874a
Hsd17b14
Slc9a3r2
Nkg7
Spag4
Cdc14b
Pcbd1
Arv1
Engase
Dusp18
Nrxn1
Srr
Gm15658
Zxdc
Bhlha15
Cela1
Amacr
Fkbp11
Gm15581
Sipa1l1
Ttll1
Wdr60
Ccdc66
Dok4
Dusp4
Afap1l1
Igkv12-89
B9d1
Nlrx1
Fkbpl
Serpinb1a
Thap2
Zfp503
Zfp184
Rasd1
4933404O12Rik
Zfp329
Egln3
Ttll9
Gm11998
Gm43489
Rbpms
Lamc1
Polr3g
Ankrd26
Ap4e1
Ino80dos
Arap2
Usp31
Ypel1
Dzip3
Rdh13
Pofut1
Prodh
Zfp428
Zscan22
Spata7
Lyrm9
Ppm1f
Amot
Cnn3
Anxa9
Msrb2
Gm5530
Rptor
Golim4
Dnajb4
Gm16897
Slc17a9
Fancf
Calcrl
Ero1l
Myl4
Cbr1
Brwd3
Hist1h3d
Kdm5b
Mphosph9
Polr3d
Meis1
Zfp703
Gm17936
Rttn
Iqsec1
Cep57l1
Ldlrad3
Cep89
Gprasp1
Snapc4
Ttll5
D3Ertd751e
Bsn
Dgat2
Wwc2
Cmtr2
Plscr1
Ptch1
Hoxa6
CD81
Elk3
Tsen15
Ppfia4
Atat1
Shmt1
Hoxa5
Gm6505
1810014B01Rik
Ovca2
Ydjc
Psat1
Mmd
Zfp799
Mccc2
Klf16
Pcp4l1
Rmnd5a
Pitpnc1
Gm7512
Smyd5
Gm4879
Gm20156
Sumf2
Ifi27
Gm9826
Dtnbp1
Fasn
Ccl9
Nt5dc2
Hspd1-ps3
Msi2
Pfkl
Fam65a
Abhd2
PdCD4
Nfe2
Mthfd1l
Pabpc4
Gm23130
Nipsnap3b
Zbed4
Tfb2m
Lyrm2
1810032O08Rik
Gm5251
3110040N11Rik
Mtap
Susd1
Sigirr
Hspd1
Hibch
Ppif
Gm16465
Gm9260
Gm6166
Dkc1
Pop1
Pcyox1l
HaCD1
2610318N02Rik
Snora68
Timm9
Ruvbl2
Cdca7l
Sos2
Bora
Cycs
Prps1
Gm17230
Rpl21
2610002M06Rik
Ndufa4
Mif-ps3
Trip13
Stap1
Gm4258
Mt1
Lmo2
Mkrn2
Gm6960
Crtap
Phb
Sh2d5
Hk3
Hoxa7
Ssfa2
Ncl
Nudcd2
Ptcd3
Adsl
Gm4705
Polr2f
Gsr
Cisd3
Clec4d
Ttc3
Ppan
Rpl12
Gm9755
Tmem97
Ddx18
Gm21887
Stk16
Gm13937
mt-Te
Psmd5
Gm7289
Cnpy4
Rsl1d1
Syncrip
Gspt1
Ssbp1
Gm8520
Arid4b
Gm5297
Gm15654
Cnpy2
Gm7389
Syce2
Dnajc2
Atp5g1
Rps8
Gm27219
Eif3j1
Nop2
Paip1
Dtymk
Gm12583
Sfxn1
Noc2l
Hdgf
Golgb1
Gm13776
Cbx3
Rps13-ps4
Rsbn1l
Hspbp1
Cct3
Psmg1
Cand1
Gm5687
Nop10
Gm19031
Ndufb2
Hmgn5
Ipo5
Fto
Zbed3
Msh6
Rcn2
Pfas
C1qbp
Gm42814
Gm11993
Gm8459
Pgam1
Gm13498
Gm15770
Ift27
Eno1b
Mcm6

Slc29a2
Hemk1
Tnfrsf22
Zfp160
Spa17
Ascl2
Pif1
Fam188b
Gm37060
Fktn
Raet1d
Gpam
Ctsg
Rsph9
Lekr1
2810403D21Rik
Itih5
9230114K14Rik
Armcx6
Sntb2
Raet1e
Nrg1
Per3
Zfp113
Mex3a
Tom1l1
Ttc26
Gm1976
Zfpm1
Tesc
Rbp1
Gmpr
Rab44
Col11a2
Acot1
Ppargc1b
Zc3hav1l
Asap2
Ccdc125
SpoCD1
Irx2
Jade2
Chsy1
Rhoq
Epha4
Kcnh7
Dus4l
Zcwpw1
Vkorc1l1
Tctn3
Mgarp
Fbxo25
4930520O04Rik
1700113H08Rik
Pomgnt1
Nicn1
Pcsk4
Sgsm2
Gm20342
Mmp12
Mllt3
Zfp606
Dand5
Socs2
Myb
Hexim2
Adgrg7
Pex10
Gfod2
Zfat
Bbs12
Zfp597
Nhsl1
Snap47
Ldlrad4
Ada
Cry2
Repin1
Acyp1
Parp16
Gm14057
Dtd1
Ap3s1-ps2
1190002N15Rik
Slc19a1
Arhgap18
Aldh6a1
Cd34
Gm5874
Hist1h3f
Ccdc40
Rnf219
Tmem69
Fam212a
Cd320
Gm17149
Trp53rkb
Cenpv
Nol8
Slc25a15
Ercc6
Pde4d
Hgh1
Cdca7
Wdr34
Il12a
1110051M20Rik
Gm9797
Mcm3ap
Hilpda
Smad3
Zfp280c
Nup43
Ipp
Dhx33
Tsc22d1
Plscr3
Zfp369
Rpgrip1
Fabp5
Fam78a
Ap3s1
Sdhaf3
Cirh1a
Timm8a1
Hspa13
Vwa8
Ap3s1-ps1
Tex9
Proser3
Frat2
Osbpl5
Tfpi
Irak1bp1
St14
Serf1
Mmachc
Tmem186
Papss2
Gm4535
Scly
Ddah2
Zc3h7b
Zfp687
Miga1
Xpc
2410016O06Rik
Rfx3
Ctps
Gm11451
Zcchc11
Mtf2
Msto1
Gm5551
Hcst
Mrpl44
Gm5855
Lamp3
Ppp1cb
Dlat
Fchsd2
Nedd4
1110038B12Rik
Nutf2
Ska2l-ps
Bace1
Hsd17b10
Snhg5
Tmem107
Gm5837
Ppid
Gm12020
Asph
2810004N23Rik
Umps
Nop58
Gm14630
Fkbp4
Hells
Dusp12
Lsm3
Hist1h2ac
Pdrg1
Psmg2
Gm14681
Fam133b
Slc16a1
Dnajc10
Ak2
Gm17251
Dph5
Haus3
Gm7933
Rpl36a
Idh3a
Rasa1
Pgm2
Ppat
Etfb
Cd72
Tfdp2
Ppcdc
Gm13521
Npm1
Gm10110
Atad3a
Paip2b
Timm10
Klhl9
Rcc1
Tspan5
Chek1
Gng12
Polr1c
Gm11349
Abce1
Rcl1
Pnkd
2610524H06Rik
Gm5525
Eif3b
Akap13
Rnaseh2c
Mrps18b
Tnpo1
Polr3k
Npm3
Wdr74
Mcfd2
Mdh2
Idi1
Pa2g4
Gm11516
Chaf1b
Got2
B3gnt2
Med29
Sephs2
Fmc1
Ilf2
Nhp2l1
Mcee
Cct6a
Flt3
Calu
Gm42418
Ogt
Gm9531
Paics
Kpnb1
Wdr36
Hspa9-ps1
Pcyt2
Fxn
Mlec
Prdx6
Gm14857
mt-Tt
Dars
Brix1
Snhg6
1810022K09Rik
Yaf2
Tcp1-ps1
Prdx3
Ube3a
Rrn3
Nudt19
Prmt5
Hsp90ab1
Twist1
Mrpl12
Npm3-ps1
Dnm1l
Gm8069
Gm8741
Gm12955
Nudt3
Gnpat
Psmd7
Thyn1
Gm7251
Prdx6b
2810025M15Rik
Dut
Rbmxl1
Gm12848
Dcun1d5
Chtf8
Gm10131
Luc7l
Mcm7

Vsig10l	Zfp146	Gm43213	Abcf2
Phc1	Alg2	Kdelc1	Cacybp
Gimap5	Gm12611	Atr	Tdg
Ell2	2300009A05Rik	Rad18	Gm6065
Elane	4930581F22Rik	Gm3362	Rpl36-ps2
Nlgn2	Nmnat3	Ern1	Nip7
P3h1	Sdc1	Gm15800	Ccnh
Ccdc88c	9330159M07Rik	Amd1	Faf1
Gm16194	Vat1	Thop1	Erh
2610307P16Rik	Hist1h2bk	Gm15528	Ddx21
Zfp335os	Gm7901	Prtn3	AI506816
Zfp707	Slc20a2	Polr3h	Scd2
Mterf2	Gm16302	Mybbp1a	Ccdc58

Table 11: 653 genes which are upregulated during differentiation of EV cells, identifying those genes which have an associated C/EBPα peak in both ChIP-seq replicate experiments at Day 5 of differentiation

1600014C10RIK	DPP4	LPP	RNF166
2310035C23RIK	DPY19L1	LPXN	RNH1
3110043O21RIK	DPYSL2	LRCH1	RNPEP
5031439G07RIK	DRAM1	LRP1	RPS6KA1
AB124611	DTX2	LRRC25	RSAD2
ABCG1	DUSP1	LRRC75A	RTP4
ABHD12	DUSP10	LRRFIP2	RUNX1
ABHD15	DUSP22	LRRK1	S100A11
ABI1	DUSP3	LSP1	SAMD8
ACAA1A	DYRK1A	LST1	SAMHD1
ACBD5	E2F2	LY6E	SAT1
ACER3	EGR1	LY86	SBNO2
ACOT9	EHD4	LY9	SCARB2
ACSS1	ELMO2	LYN	SCIMP
ACTA2	ELMSAN1	LYZ1	SDC3
ACTR3	EMILIN2	LYZ2	SDC4
ADD3	ENSA	MACF1	SEC24B
ADGRE1	ENTPD1	MALAT1	SELPLG
ADGRE4	EPB41L2	MALT1	SEMA4A
AFF1	EPB41L4B	MAML1	SEMA4D
AHNAK	ERLIN1	MAP3K1	SESN2
AI839979	ETV3	MAP3K3	SGK3
AIM2	EVL	MAP3K8	SGPL1
AKR1A1	F630028O10RIK	MAP4	SGPP1
AKT3	FAM105A	MAPK14	SH3BP2
ALDH3A2	FAM129A	36951	SH3GLB1
ALOX5	FAM129B	MAX	SIDT2
ALOX5AP	FAM174A	MBNL1	SIGLEC1
ANKRD44	FAM46C	MCL1	SIRPA
ANXA1	FAM49B	MDFIC	SKIL
ANXA2	FAM69A	MEF2A	SLAMF7
ANXA6	FAR1	METRNL	SLAMF9
AP2A2	FBRSL1	MFSD5	SLC15A3
APBA1	FCGR3	MGAT4A	SLC15A4
APBB1IP	FCGR4	MICAL1	SLC16A7
APOBEC1	FEM1C	MILR1	SLC29A3
ARHGAP15	FES	MKNK1	SLC2A6
ARHGAP25	FGD3	MOB3A	SLC31A1
ARHGAP31	FGR	MPEG1	SLC35C1
ARID5A	FILIP1L	MS4A4A	SLC37A2
ARL4C	FKBP15	MS4A4B	SLC43A2
ARL5C	FKBP1B	MS4A4C	SLC44A2
ARL6IP5	FLNA	MS4A6B	SLC6A6
ARPC5	FMNL1	MS4A6C	SLC7A8
ARRB2	FNBP1	MS4A6D	SLFN1
ASGR2	FNIP2	MSRB1	SLFN2
ATF3	FOS	MTHFS	SMIM3
ATG7	FOXN2	MYADM	SMOX
ATP2B1	FOXRED2	MYD88	SMPDL3B
ATP6V1B2	FRMD4B	MYO1C	SNX10
AXL	FTH1	MYO1F	SNX2
AZI2	FTL1	MYO9B	SNX20

AZIN1
B430306N03RIK
B4GALT5
BACH2
BC005537
BCL2L11
BCL3
BCL6
BEX6
BHLHE40
BIN1
BLVRB
BROX
BTG2
C3
C3AR1
CADM1
CAMK1D
CAP1
CAPG
CAPN2
CAPZA2
CARD11
CARD19
CASP4
CAST
CBL
CCDC109B
CCDC12
CCDC180
CCDC50
CCDC86
CCL2
CCL3
CCL6
CCR1
CCR2
CCRL2
CD14
CD22
CD244
CD28
CD300A
CD300LB
CD300LF
CD48
CD52
CD53
CD68
CD74
CD82
CD86
CD9
CDC42SE1
CDC42SE2
CDKN1A
CFH
CHMP4B
CHN2
CIB1
CKB
CLCN7
CLEC4A2
CLEC4B1
CLEC4N
CLEC7A
CLN8
CMIP
CMPK2
CMTR1
CNDP2
CNR2
CORO1A
FYB
G6PDX
GAS7
GBP2
GCNT1
GCNT2
GGH
GM1966
GM6377
GM7367
GM8369
GNA15
GNG2
GNS
GPR108
GPR132
GPR141
GPR146
GPR155
GPR160
GPR183
GPR34
GRB2
GRN
GSAP
GSDMD
GSN
HACD4
HCAR2
HCFC1R1
HCK
HDAC4
HERPUD1
HEXA
HEXB
HK2
HLX
HMHA1
HMOX1
HPCAL1
I830077J02RIK
ICOSL
ID3
IDS
IER2
IER5
IFI205
IFITM3
IFITM6
IFNAR1
IFNAR2
IFNGR1
IFNGR2
IGSF6
IL10RA
IL13RA1
IL15
IL17RA
IL18
IL1R2
IL21R
IL3RA
IL4RA
IL6RA
IL6ST
INPP4A
IQGAP1
IQGAP2
IRAK2
IRF2BP2
IRF5
IRF8
ISG20
N4BP2L1
NAGPA
NAIP5
NAPSA
NCF2
NCKIPSD
NCOR2
NDEL1
NEAT1
NECAP1
NEDD9
NEK7
NFAM1
NFATC1
NFE2L2
NFKBIA
NFKBIZ
NLRP3
NOS1AP
NOTCH2
NPC2
NPTN
NR4A1
NRP1
NRROS
NUAK2
NUB1
NXPE4
OSBPL9
P2RX4
P2RX7
P2RY6
PADI2
PAG1
PAK1
PDXK
PELI1
PFKFB4
PHACTR2
PIAS1
PID1
PIK3AP1
PIK3CD
PIK3R5
PILRA
PIP4K2A
PIRB
PISD
PLAC8
PLAUR
PLCL2
PLD4
PLEC
PLEKHA1
PLEKHF2
PLEKHG3
PLEKHM1
PLEKHM3
PLEKHO2
PLIN2
PLXDC1
PLXNC1
PMAIP1
PNKP
PNPLA7
POT1B
PPP2R5A
PRDX5
PREX1
PRKAB2
PRKCD
PRKCH
PRKD3
SOAT1
SOCS1
SOCS3
SORL1
SP100
SP2
SPG21
SPI1
SPOP
SPTY2D1
SQSTM1
SRGAP2
SSH1
ST18
ST3GAL2
ST8SIA1
ST8SIA4
STAT1
STAT6
STK17B
STX7
STXBP3
SUB1
SUSD3
SYK
SZRD1
TAGAP
TALDO1
TBC1D13
TBC1D8
TBC1D9
TDRD7
TEC
TECPR1
TEP1
TET2
TFEB
TGFBI
THAP6
THEMIS2
TIFA
TIPARP
TLE3
TLR1
TLR13
TLR2
TLR4
TLR6
TMEM106A
TMEM150B
TMEM154
TMEM184B
TMEM189
TMEM192
TMEM229B
TMEM243
TMEM50B
TMOD3
TMSB10
TMSB4X
TNFAIP2
TNFRSF13B
TNFRSF14
TNFRSF1A
TNFSF8
TOR1A
TOR1AIP1
TOR3A
TPD52
TRAFD1
TREM2
TSPYL1
TTC7

CORO1B	ITGA1	PRKX	TUBGCP5
COTL1	ITGAM	PRR13	TULP4
CRLF2	ITGAX	PSAP	TXNDC16
CTNNA1	ITGB2	PSEN1	TYROBP
CTNND2	ITM2B	PSTPIP1	UBC
CTSB	ITPR1	PSTPIP2	USP12
CTSC	JAK2	PTAFR	USP18
CTSD	JUNB	PTGER4	USP25
CTSE	KDM7A	PTK2B	UVRAG
CTSZ	KIF1B	PTPN1	VAPB
CTTNBP2NL	KLF10	PTPN22	VASP
CX3CR1	KLF13	PTPRA	VCL
CXCL16	KLF3	PTPRC	VIM
CXCL2	KLF4	PTPRJ	VPS37B
CXCR4	KLF6	RAB3D	VRK1
CYB561A3	KLHL6	RAB8A	VSIR
CYBA	KLRA17	RACGAP1	WDR1
CYBB	KLRB1F	RALA	WDR26
CYFIP1	KLRK1	RAP1B	WHSC1L1
CYP4F16	LACC1	RAP2C	WIPF1
CYP4F18	LACTB	RAPH1	XBP1
CYTH4	LAIR1	RARA	YTHDF3
CYTIP	LAMP1	RARG	ZBP1
D1ERTD622E	LAPTM4A	RASGEF1B	ZBTB7A
DAB2	LASP1	RASSF4	ZBTB7B
DBNL	LBH	RASSF5	ZC3H12A
DCAF11	LCP1	RBM47	ZC3HAV1
DCK	LCP2	RBMS1	ZCCHC2
DCLRE1C	LGALS3	RCBTB2	ZDHHC7
DDIT3	LGMN	RCHY1	ZFHX3
DEGS1	LIFR	RELL1	ZFP263
DENND1A	LILRA6	RGL1	ZFP277
DENND1B	LILRB4A	RGS2	ZFP36L2
DHX58	LITAF	RGS3	ZFP710
DIAPH2	LMLN	RHOG	ZFYVE26
DNA2	LMNA	RIN2	ZYX
DNAJB14	LPCAT4	RIN3	
DOCK4	LPIN1	RMND5B	
DPEP2	LPL	RNF115	

Table 12: Subset of 230 genes from Table 11 which are also upregulated in *CEBPA* N321D-transduced cells during differentiation in Flt3L

5031439G07RIK	FAM105A	LST1	RMND5B
ABHD15	FAM129B	LY6E	RPS6KA1
ABI1	FAM174A	LY86	RTP4
ACAA1A	FES	LYN	RUNX1
ACER3	FILIP1L	LYZ2	SAMHD1
AHNAK	FKBP15	MAP3K1	SAT1
AI839979	FMNL1	MAP3K8	SDC3
ALOX5AP	FRMD4B	MAP4	SDC4
ANXA6	FYB	36951	SEMA4D
APBA1	GAS7	MBNL1	SLC16A7
ARL4C	GCNT1	MEF2A	SLC2A6
ARL6IP5	GCNT2	MILR1	SLC44A2
AXL	GNG2	MPEG1	SLC6A6
BACH2	GNS	MS4A4B	SLFN2
BCL6	GPR141	MS4A4C	SMIM3
BEX6	GPR183	MS4A6B	SNX20
BLVRB	GPR34	MS4A6C	SOAT1
CADM1	GRN	MS4A6D	SP100
CAMK1D	GSN	MYADM	SPI1
CAPG	HLX	MYO1F	ST8SIA4
CCDC180	HMHA1	NAIP5	STAT1
CCL3	I830077J02RIK	NAPSA	STK17B
CCL6	IFI205	NEAT1	STX7
CCR1	IFITM3	NECAP1	TAGAP
CD14	IFITM6	NFAM1	TDRD7
CD28	IFNAR2	NFATC1	TEC
CD300A	IL10RA	NFKBIA	TECPR1
CD300LB	IL13RA1	NLRP3	TEP1

CD300LF	IL17RA	NOS1AP	TET2
CD52	IL1R2	NOTCH2	TFEB
CD68	IL6RA	NPTN	THEMIS2
CD74	IL6ST	NR4A1	TIFA
CD82	INPP4A	NRROS	TIPARP
CDC42SE1	IRF2BP2	P2RX7	TLR2
CDKN1A	IRF5	P2RY6	TMEM229B
CFH	IRF8	PAK1	TMEM243
CKB	ITGA1	PIK3R5	TMEM50B
CLEC4A2	ITGAM	PILRA	TMSB10
CLEC7A	ITGAX	PIP4K2A	TNFRSF13B
CLN8	ITGB2	PIRB	TOR1AIP1
CMPK2	KDM7A	PLAC8	TOR3A
CTNNA1	KLF13	PLD4	TPD52
CTNND2	KLF3	PLEC	TRAFD1
CTSB	KLHL6	PLEKHA1	TUBGCP5
CTSE	KLRB1F	PLEKHF2	TULP4
CTSZ	KLRK1	PMAIP1	TYROBP
CX3CR1	LAIR1	PREX1	USP12
CYBB	LAMP1	PRKCD	USP18
CYFIP1	LASP1	PRKX	VSIR
CYTH4	LBH	PRR13	WDR26
CYTIP	LCP2	PSAP	WHSC1L1
D1ERTD622E	LGALS3	PSTPIP2	YTHDF3
DHX58	LGMN	PTPRC	ZBTB7A
DPEP2	LIFR	RARA	ZC3HAV1
DUSP22	LILRB4A	RASSF4	ZFP263
EHD4	LITAF	RASSF5	ZFP36L2
EPB41L4B	LPXN	RBM47	
EVL	LSP1	RIN2	

Table 13: Subset of 126 genes from Table 12 which are also bound by FLAG antibody at Day 5

ACER3	FRMD4B	LY86	RASSF4
AHNAK	FYB	LYN	RBM47
AI839979	GAS7	MAP3K1	RIN2
APBA1	GCNT1	MAP4	RPS6KA1
AXL	GCNT2	36951	RTP4
BLVRB	GNG2	MBNL1	RUNX1
CADM1	GNS	MEF2A	SDC4
CAMK1D	HLX	MILR1	SEMA4D
CCL3	I830077J02RIK	MPEG1	SLC16A7
CCL6	IFITM3	MS4A6B	SLC2A6
CCR1	IFITM6	MYO1F	SMIM3
CD14	IFNAR2	NAIP5	SOAT1
CD28	IL17RA	NECAP1	SPI1
CD300A	IL6RA	NFAM1	ST8SIA4
CD300LB	IL6ST	NFATC1	STK17B
CDKN1A	INPP4A	NLRP3	STX7
CLEC7A	IRF2BP2	NOS1AP	TAGAP
CMPK2	IRF8	NOTCH2	TDRD7
CTSB	ITGAM	NR4A1	TEC
CTSE	ITGAX	NRROS	TEP1
CTSZ	KLF13	PILRA	TET2
CYFIP1	KLF3	PIP4K2A	THEMIS2
CYTH4	KLHL6	PLAC8	TIPARP
D1ERTD622E	KLRB1F	PLD4	TLR2
DPEP2	LAIR1	PLEKHA1	TMEM229B
DUSP22	LAMP1	PLEKHF2	TMEM243
EHD4	LBH	PMAIP1	TPD52
EPB41L4B	LCP2	PREX1	WHSC1L1
FAM105A	LGALS3	PRKCD	ZC3HAV1
FAM174A	LGMN	PSAP	ZFP263
FES	LITAF	PSTPIP2	
FKBP15	LY6E	PTPRC	

Table 14: Subset of 36 genes from Table 13 which are downregulated on differential expression between *CEBPA* N321D versus EV-transduced cells at Day 5

ACER3	DPEP2	LY86	PREX1
AHNAK	FYB	MPEG1	PRKCD
AXL	IFITM3	MS4A6B	PSAP
CCL6	IFITM6	MYO1F	RTP4
CCR1	IRF2BP2	NFAM1	TDRD7
CD300A	ITGAM	NOTCH2	THEMIS2
CMPK2	LGALS3	NR4A1	TLR2
CTSB	LITAF	PLD4	TPD52
CTSZ	LY6E	PLEKHF2	ZC3HAV1

Table 15: Genes which have an associated FLAG peak at Day 5 in both replicate ChIP-seq experiments

160 genes with high-confidence FLAG peak at Day 5 which are also downregulated on differential expression between *CEBPA* N321D and EV-transduced cells at Day 5

ABCG1	DCK	LACC1	PSTPIP1
ABHD12	DEGS1	LAPTM4A	PTAFR
ACER3	DENND1B	LGALS3	PTPN1
ACTA2	DPEP2	LITAF	RAP1B
AHNAK	DPP4	LMLN	RBPMS
ALPK1	DUSP1	LPCAT2	RCBTB2
AMZ1	EMILIN2	LPCAT4	RGS2
ANXA2	EP300	LY6E	RHOG
AOAH	ESYT1	LY86	RSAD2
APOBEC1	FAM129A	LYZ1	RTP4
ATF3	FAM69A	MAN1A	SEC24B
ATP10A	FCGR4	MAPK14	SIK3
ATP2B1	FOS	MARCH1	SIRPA
ATP6V1B2	FYB	MCTP1	SKIL
AW112010	GBP2	MPEG1	SLC15A3
AXL	GM1966	MS4A6B	SLC7A11
B4GALT5	GPR160	MS4A7	SLC7A8
C3AR1	GSAP	MSRB1	SMPDL3B
CCL6	HACD4	MYCBP2	SORL1
CCR1	HCAR2	MYO1F	ST8SIA1
CD180	HIST1H1C	NABP1	STK38
CD22	HMOX2	NAMPT	TBC1D8
CD300A	ID3	NFAM1	TDRD7
CD84	IER2	NOTCH2	TGFBI
CD9	IFITM1	NPC2	THEMIS2
CDK14	IFITM3	NR4A1	TLR1
CLEC10A	IFITM6	NXPE4	TLR13
CLEC4A1	IFNGR2	PDE7B	TLR2
CLEC4N	IL15	PID1	TMEM150B
CLEC5A	INF2	PIK3AP1	TMSB4X
CMPK2	IQGAP1	PIM1	TNFRSF1A
CNDP2	IRAK2	PLD4	TNFSF8
COTL1	IRF2BP2	PLEKHF2	TPD52
CSF3R	ITGAM	PLEKHM3	TTC7
CTSB	ITM2B	PLIN2	WIPF1
CTSC	ITPR1	PPP2R5A	XDH
CTSZ	JAK2	PREX1	ZBP1
CTTNBP2NL	KCNK13	PRKCD	ZC3H12A
CYP4F18	KLF6	PRKCH	ZC3HAV1
DAB2	L1CAM	PSAP	ZYX

49 genes with high-confidence FLAG peak at Day 5 which are also upregulated on differential expression between *CEBPA* N321D and EV-transduced cells at Day 5

8430429K09RIK	COL19A1	INSL6	SLC44A1
ABCD2	CTR9	KATNAL1	SLPI
ABHD6	CTSG	LDLRAD4	SOCS2
AFAP1	D930015E06RIK	MYB	SORCS2
AK2	DHRS3	MYOG	TACO1
ARAP2	DR1	NEIL1	TMEM229B

ATP1B3
AW209491
B3GNT2
BACE1
CBX2
CHSY1
COBLL1
DTX4
EPHA7
FNDC3B
GAS2
HTRA3
IL6ST
INO80DOS
PI15
RAB44
RCN1
SDC1
SERPINB1A
SGCG
SIX1
TRP53RKB
XPC
ZFP369
ZNRF1